

UNIVERZITA OBRANY
FAKULTA VOJENSKÝCH TECHNOLOGIÍ
KATEDRA MATEMATIKY A FYZIKY

NUMERICKÉ METODY

JAROMÍR KUBEN
PAVLÍNA RAČKOVÁ

BRNO 2019

Kuben Jaromír, Račková Pavlína
Numerické metody

© Jaromír Kuben, Pavlína Račková 2019

ISBN 978-80-7582-092-1

Předmluva

Tento elektronický e-learningový text obsahuje základní poznatky z numerických metod a je primárně určen studentům FVT UO, Brno. Lze jej však využít při výuce základů numerických metod na jakýchkoli vysokých školách, zejména technického zaměření.

Numerické metody jsou na FVT UO součástí předmětů v několika studijních programech. Podle toho se liší probíraný rozsah látky. Studentům tříletého bakalářského programu Technologie pro obranu a bezpečnost a končícího pětiletého magisterského programu (SP 2014) Vojenské technologie jsou určeny kapitoly 1–6, zatímco studentům dvouletého nadstavbového magisterského programu Vojenské technologie jsou určeny kapitoly 1, 2, 4, 6 a 7 (s tematikou kapitol 3 a 5 se seznámili již v předcházejícím studiu). V nových pětiletých magisterských programech (SP 2019) Vojenské technologie strojní resp. elektrotechnické budou využity kapitoly 1–7. Některé kapitoly obsahují náročnější rozšiřující materiál, který není určen pro výuku ve zmíněných základních kurzech, ale bude využit v předmětu Numerická matematika v doktorských studijních programech.

Numerické metody patří bezesporu k základnímu matematickému vzdělání inženýra. Potřeba řešit neustále složitější matematické modely a existence čím dál výkonnějších počítačů vede k bouřlivému rozvoji numerické matematiky. Bohužel mnohé soudobé metody jsou velmi složité a jejich vysvětlení nematematikovi je velice obtížné až nemožné. Nicméně většina těchto metod staví na klasických poznatcích z numerické matematiky. Cílem předkládaného textu je předvést průřez nejdůležitějšími partiemi numerické matematiky a seznámit studenty se základními klasickými výsledky.

Pro správné pochopení numerických metod je bezpodmínečně nutné si je „ohmatat“ a prakticky vyzkoušet na počítači. Jen pak může jejich výuka přinést potenciálním uživatelům nějaký užitek. Téměř každý student na vysoké škole technického zaměření se totiž časem setká s potřebou vyřešit nějakou matematickou úlohu nemající řešení dané explicitním vzorcem. Ve skutečnosti je většina reálných a z praktického hlediska důležitých a užitečných úloh tohoto druhu. Takovou úlohu je nutné řešit přibližně pomocí vhodné numerické metody. Studenti by měli být na takovou situaci připraveni a měli by vědět, kam sáhnout. Proto je většina cvičení prováděna na počítači s vhodným programovým vybavením. Existuje celá řada komerčních i volně šiřitelných velmi kvalitních programů, které poskytují potřebné nástroje. Z nich lze jmenovat např. komerční programy Maple, Matlab, Mathematica, Mathcad a volně šiřitelné programy Maxima, Sage, Octave a řadu dalších.

Jelikož počet hodin na výuku numerické matematiky je dost omezený, museli jsme pomínout některé podstatné partie (např. partie z lineární algebry jako vlastní čísla a vektory matic a řadu dalších, optimalizaci, okrajové úlohy pro řešení obyčejných diferenciálních rovnic, variační metody pro řešení parciálních diferenciálních rovnic, metodu konečných prvků apod.). Zájemci o ně najdou odkazy v seznamu literatury. Protože text je určen především inženýrům, je naprostá většina poznatků uvedena bez důkazů. Čtenář, který by se s nimi chtěl seznámit, najde v textu četné odkazy na vhodnou literaturu.

Skriptum obsahuje poměrně rozsáhlý seznam literatury od elementárních textů až po specializované monografie. V současnosti je dostupná řada velmi pěkných úvodních i pokročilejších textů, jak v tištěné podobě tak v elektronické podobě na Internetu. Z úvodních textů lze doporučit např. [9, 11, 13, 41], z pokročilejších pak [8, 22, 52]. V seznamu lze nalézt většinu podstatných publikací z numerické matematiky, které byly česky celostátně vydány za posledních 50 let.

Základem tohoto materiálu byl učební text [35], který byl obohacen o prvky charakteristické pro e-learning. Do každé kapitoly byly doplněny cíle, pojmy k zapamatování, kontrolní otázky, cvičení s výsledky (pokud byla vhodná pro ruční výpočty) a interaktivní testy. Za poslední kapitolu byly přidány souhrnné interaktivní testy, pokrývající všechny kapitoly. Celý materiál je hypertextový, obsahuje řadu odkazů jak v rámci textu (vzorce,

věty, definice, rejstřík, obsah apod.), tak na internet (např. údaje o významných matematicích, jejichž jména se v textu objevují, nebo na studijní zdroje). Všechny ilustrace jsou nyní barevné, některé statické obrázky byly nahrazeny animacemi resp. pohyblivými 3D objekty. Po obsahové stránce byl text rozšířen. Zejména byly zařazeny ukázkové příklady v kapitole 2. Dále byl v kapitole 4 rozšířen oddíl o Hermitově interpolačním polynomu a v kapitole 5 přidány oddíly o numerické derivaci a její podmíněnosti.

Děkujeme recenzentům doc. RNDr. Liboru Čermákovi, CSc. z Ústavu matematiky Fakulty strojního inženýrství VUT v Brně a Mgr. Jiřímu Zelinkovi, Dr. z Ústavu matematiky a statistiky Přírodovědecké fakulty Masarykovy univerzity v Brně za velmi cenné připomínky, které přispěly ke zkvalitnění textu. Rovněž děkujeme kolegům z naší katedry Mgr. Vojtěchu Růžičkovi, Ph.D. a Mgr. Jaromíru Kubenovi za pečlivou kontrolu interaktivních testů a druhému jmenovanému rovněž za důkladné přečtení celého textu, připomínky k němu a zejména za větu 4.13.

Text ve formátu PDF byl připraven sázecím systémem $\text{T}_{\text{E}}\text{X}$ pomocí $\text{T}_{\text{E}}\text{X}$ ového formátu $\text{pdfL}_{\text{A}}\text{T}_{\text{E}}\text{X} 2_{\epsilon}$, statické obrázky byly vytvořeny programem METAPOST s použitím balíku $\text{T}_{\text{E}}\text{X}$ ovských maker `mfpic`, animace a 3D obrázky byly připraveny v programu Maple. Interaktivní testy byly vytvořeny s použitím $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ ovského balíku Acrotex. Aby animace, 3D obrázky a testy fungovaly, je nutné použít jako prohlížeč Adobe Reader (nebo plný

Adobe Acrobat). Dále je třeba v menu Úpravy/Předvolby/3D a multimédia zatrhnout volbu *Aktivovat přehrávání 3D obsahu* a vhodné je rovněž zatrhnout volbu *Povolit oboustranné vykreslení*. V jiných prohlížečích tyto prvky nebudou funkční. Protože na některých počítačích po otevření určitých 3D objektů dochází v důsledku nedostatečných parametrů grafiky k „zamrzávání“ Adobe Readeru, je u všech 3D objektů nejprve znázorněn bitmapový náhled (ten je vidět i v jiných prohlížečích) a k aktivaci dojde až po kliknutí na jeho plochu (pokud je povoleno přehrávání 3D objektů). Problémy nejsou v plném Adobe Acrobatu.

Brno, červen 2019

autoři

Obsah

Předmluva	3
1 Úvod do problematiky numerických metod	16
1.1 Zdroje chyb	19
1.2 Aproximace čísel	20
1.3 Reprezentace čísel v počítači	26
1.4 Korektnost a podmíněnost úlohy	29
1.5 Vlastnosti numerických algoritmů	34
1.6 Symbol O	36
1.7 Vektorové a maticové normy, skalární součin	39
1.7.1 Vektorové normy	40
1.7.2 Maticové normy	49
1.7.3 Číslo podmíněnosti čtvercové matice	64

1.7.4	Spektrální poloměr	67
1.7.5	Skalární součin	69
	Pojmy k zapamatování	86
	Kontrolní otázky	87
	Cvičení	89
	Řešení	94
	Testy ke kapitole 1	97
	Test 1	97
	Test 2	101
	Test 3	105
	Test 4	109
2	Numerické řešení nelineárních rovnic	113
2.1	Nelineární rovnice o jedné neznámé	114
2.1.1	Postup při hledání kořenů	116
2.1.2	Zastavovací podmínky	119
2.1.3	Rychlost konvergence	121
2.2	Iterační metody pro řešení rovnic o jedné neznámé	123
2.2.1	Klasifikace iteračních metod	123

2.2.2	Metoda bisekce (metoda půlení intervalu)	125
2.2.3	Metoda regula falsi (metoda tětív)	131
2.2.4	Metoda prosté iterace	143
2.2.5	Newtonova metoda (metoda tečen)	154
2.2.6	Metoda sečen	161
2.3	Systémy nelineárních rovnic	170
2.3.1	Metoda prosté iterace	173
2.3.2	Newtonova metoda pro systémy	175
	Pojmy k zapamatování	181
	Kontrolní otázky	183
	Testy ke kapitole 2	187
	Test 1	187
	Test 2	193
	Test 3	198
3	Numerické řešení systémů lineárních rovnic	203
3.1	Přímé metody	207
3.1.1	Gaussova eliminační metoda	207
3.1.2	Metoda LU rozkladu	229

3.1.3	Choleského metoda	244
3.1.4	Analýza chyb při numerickém řešení soustav lineárních rovnic	249
3.2	Iterační metody	257
3.2.1	Jacobiova metoda	263
3.2.2	Gaussova-Seidelova metoda	268
3.2.3	Relaxační metody, metoda SOR	275
	Pojmy k zapamatování	280
	Kontrolní otázky	282
	Testy ke kapitole 3	287
	Test 1	287
	Test 2	292
	Test 3	297
4	Interpolace a aproximace funkcí	303
4.1	Interpolační polynom	307
4.1.1	Lagrangeův tvar interpolačního polynomu	309
4.1.2	Newtonův tvar interpolačního polynomu	317
4.1.3	Newtonův interpolační polynom — ekvidistantní uzly	324
4.1.4	Hermitův interpolační polynom	335

4.2	Interpolace splajny	366
4.3	Aproximace metodou nejmenších čtverců	389
4.3.1	Obecný případ	389
4.3.2	Vyrovnnání polynomy	396
4.3.3	Určení stupně aproximačního polynomu	405
	Pojmy k zapamatování	410
	Kontrolní otázky	411
	Cvičení	416
	Řešení	427
	Testy ke kapitole 4	434
	Test 1	434
	Test 2	440
	Test 3	447
	Test 4	454
5	Numerická derivace a integrace	460
5.1	Princip numerické integrace	465
5.2	Kvadrurní formule	467
5.3	Newtonovy-Cotesovy kvadrurní formule	475

5.3.1	Uzavřené Newtonovy-Cotesovy kvadrurní formule	476
5.3.2	Otevřené Newtonovy-Cotesovy kvadrurní formule	484
5.3.3	Složené Newtonovy-Cotesovy kvadrurní formule	494
5.4	Gaussovy kvadrurní formule	503
5.5	Rombergova kvadratura	517
5.6	Numerická derivace	526
5.7	Podmíněnost numerické derivace a integrace	544
	Pojmy k zapamatování	552
	Kontrolní otázky	553
	Testy ke kapitole 5	555
	Test 1	555
	Test 2	562
6	Numerické řešení obyčejných diferenciálních rovnic	568
6.1	Obyčejné diferenciální rovnice 1. řádu	570
6.1.1	Základní vlastnosti obyčejných diferenciálních rovnic 1. řádu . .	570
6.1.2	Numerické řešení diferenciálních rovnic 1. řádu	580
6.1.3	Jednokrokové metody	588
6.1.4	Vlastnosti jednokrokových metod	595

6.1.5	Metody Rungeho-Kutty	606
6.1.6	Více krokové metody	623
6.1.7	Vlastnosti více krokových metod	626
6.1.8	Příklady lineárních více krokových metod	633
6.1.9	Tuhé problémy	640
6.2	Systémy obyčejných diferenciálních rovnic	646
6.3	Diferenciální rovnice vyšších řádů	652
	Pojmy k zapamatování	661
	Kontrolní otázky	663
	Testy ke kapitole 6	666
	Test 1	666
	Test 2	671
	Test 3	677
	Test 4	682
7	Numerické řešení parciálních diferenciálních rovnic	687
7.1	Klasifikace metod	688
7.2	Metoda sítí	691
7.2.1	Princip metody sítí	691

7.2.2	Korektnost, stabilita, aproximace a konvergence	695
7.2.3	Náhrada derivací diferencemi	698
7.3	Dirichletova úloha pro Poissonovu rovnici	701
7.4	Smíšená úloha pro rovnici vedení tepla	712
7.5	Smíšená úloha pro vlnovou rovnici	729
	Pojmy k zapamatování	751
	Kontrolní otázky	753
	Testy ke kapitole 7	756
	Test 1	756
	Test 2	763
	Test 3	769
Souhrnné testy		775
	Test 1	775
	Test 2	782
	Test 3	788
Literatura		794
Rejstřík		807

Kapitola 1

Úvod do problematiky numerických metod

Cíle

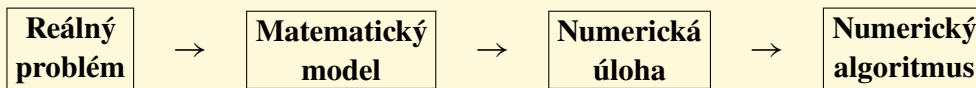
Po prostudování této kapitoly budete schopni vysvětlit:

- čím se zabývá numerická matematika,
- jaké druhy chyb vznikají při numerickém řešení úloh,
- jak jsou uložena čísla v počítači a jak se správně zaokrouhluje,
- co jsou to korektní a dobře podmíněné úlohy,
- jaké vlastnosti musí mít numerické algoritmy,
- co jsou to vektorové a maticové normy a skalární součin.

Numerická matematika se zabývá procesy, které umožňují řešit matematické problémy pomocí čtyř aritmetických operací (sečítání, odečítání, násobení a dělení) s využitím počítače, popř. s pomocí kalkulačky nebo ručně. Cílem je vytvořit efektivní algoritmy pro řešení nejrůznějších matematických problémů.

Formulace úloh a způsob jejich řešení je v dnešní době závislý na skutečnosti, že pracujeme s počítačem. To vyžaduje, abychom zadali do počítače konečný počet číselných údajů a postup, tzv. algoritmus, prostřednictvím kterého po konečném počtu kroků dostaneme na výstupu výsledek.

V praxi probíhá celý postup obvykle podle následujícího diagramu:



Všimneme si nyní jednotlivých kroků.

Reálný problém

Jde o zadanou úlohu z nejrůznějších oborů (fyzika, technika, chemie, biologie, lékařství, humanitní obory apod.), kterou chceme řešit pomocí matematických prostředků.

Matematický model

Zkoumanou skutečnost popíšeme pomocí vhodného věrohodného matematického modelu. Obvykle čím je model přesnější, tím je složitější.

Takovým modelem může být např. systém lineárních nebo nelineárních rovnic, obyčejné diferenciální rovnice, parciální diferenciální rovnice, integrální rovnice atd.

Numerická úloha

Matematický model nelze většinou řešit přímo pomocí aritmetických operací, je nutné provést jeho „digitalizaci“. Tak dostaneme úlohu, kterou lze řešit pomocí aritmetických operací. Může to být např. soustava lineárních algebraických rovnic. Vstupem i výstupem numerické úlohy jsou čísla.

Numerický algoritmus

Numerickou úlohu lze řešit různými způsoby. Numerickým algoritmem rozumíme postup řešení numerické úlohy.

Přesný popis kroků (akcí, které může realizovat počítač), které vedou k vyřešení numerické úlohy, nazýváme *numerický algoritmus*. Je to tedy jednoznačný funkční popis vztahů mezi konečným počtem vstupních a konečným počtem výstupních hodnot.

1.1 Zdroje chyb

Při řešení reálných problémů téměř nikdy nezískáme přesná řešení, musíme se spokojit s přibližnými řešeními, která jsou zatížena chybami. Jedním z nejdůležitějších úkolů je zorganizovat postup výpočtů tak, aby celková chyba byla co nejmenší.

Numerická řešení problémů jsou obvykle zatížena chybami (nepřesnostmi), které vznikají ve dvou oblastech: těmi, které jsou obsaženy v *matematické formulaci problému* (včetně chyb ve vstupních údajích), a těmi, které jsou způsobeny *hledáním numerickou cestou*. Do první skupiny patří:

1. *Chyby matematického modelu* — je jen „aproximací“, idealizací reálné skutečnosti.
2. *Chyby ve vstupních datech* — např. chyby v empirických hodnotách získaných měřeními nebo nepřesnosti fyzikálních konstant.

Do druhé skupiny patří:

1. *Chyba numerické úlohy* — neřeší se problém, který byl původně zadán, ale nějaká jeho aproximace (numerická úloha). Řešení numerické úlohy, které stejně většinou nejsme schopni získat zcela přesně, je pouze přibližným řešením matematického modelu.
2. *Zaokrouhlovací chyby* — jsou dvojího druhu: chyby, které vzniknou zaokrouhlením vstupních hodnot, a chyby, vznikající při aritmetických operacích na počítači. Důvodem je, že počítač pracuje pouze s konečnou množinou čísel — podrobněji viz odstavec 1.3.

1.2 Aproximace čísel

Ve výpočtech jsme často nuceni nahradit přesné číslo x přibližnou hodnotou \tilde{x} . Číslo \tilde{x} se pak nazývá *aproximací čísla x* . Rozdíl $x - \tilde{x} = \Delta x$ nazýváme *absolutní chybou*

aproximace \tilde{x} a podíl

$$\frac{x - \tilde{x}}{x} = \frac{\Delta x}{x}, \quad x \neq 0,$$

nazýváme *relativní chybou aproximace* \tilde{x} .

Jestliže pro nějaké $\varepsilon > 0$ platí $|\Delta x| \leq \varepsilon$, říkáme, že ε je *odhadem absolutní chyby*. Pak platí $\tilde{x} - \varepsilon \leq x \leq \tilde{x} + \varepsilon$. Používá se symbolický zápis $x = \tilde{x} \pm \varepsilon$.

Podobně jestliže pro nějaké $\delta > 0$ platí $|\Delta x/x| \leq \delta$, říkáme, že δ je *odhadem relativní chyby*. Pak platí $x - \delta|x| \leq \tilde{x} \leq x + \delta|x|$. Používá se symbolický zápis $\tilde{x} = x(1 \pm \delta)$.

Každé reálné číslo lze zapsat pomocí konečného nebo nekonečného dekadického rozvoje. Každé číslo má buď jeden nebo dva takové rozvoje. Má-li některé číslo dva různé rozvoje, pak jeden je konečný (od jistého místa jsou cifry nulové) a druhý je nekonečný a má od jistého místa samé devítky. Např. zápisy 23,42 a $23,419\,999\,999\dots = 23,41\overline{9}$ určují jedno a totéž číslo.

Označme d_1, d_2, \dots , kde $d_i \in \{0, 1, \dots, 9\}$, $i = 1, 2, \dots$, $d_1 \neq 0$, cifry dekadického rozvoje aproximace \tilde{x} . Nechť d_1 (první nenulová cifra zleva) stojí u mocniny 10^e ,

kde e je celé číslo. Pak

$$\tilde{x} = \pm [d_1 \cdot 10^e + d_2 \cdot 10^{e-1} + \dots + d_k \cdot 10^{e+1-k} + d_{k+1} \cdot 10^{e-k} + \dots].$$

Řekneme, že k -tá dekadická cifra d_k aproximace \tilde{x} je *platná*, jestliže

$$|x - \tilde{x}| \leq 5 \cdot 10^{e-k}, \quad (1.1)$$

tj. když se \tilde{x} liší od x nejvýše o pět jednotek řádu příslušejícího následující cifře. Pokud nerovnost (1.1) platí pro $k \leq p$, ale už neplatí pro $k = p + 1$, říkáme, že \tilde{x} má p platných cifer. Číslo

$$\pm [d_1 \cdot 10^e + d_2 \cdot 10^{e-1} + \dots + d_p \cdot 10^{e+1-p}] = \pm d_1 d_2 \dots d_p \cdot 10^{e+1-p}$$

se pak nazývá *správně zaokrouhlenou hodnotou* čísla x .

Z uvedené definice je zřejmé, že když d_k je platná cifra, jsou i všechny předcházející cifry, tj. $d_{k-1}, d_{k-2}, \dots, d_1$, také platné.

Řekneme, že aproximace \tilde{x} čísla x má k -té *desetinné místo platné*, jestliže

$$|x - \tilde{x}| \leq 5 \cdot 10^{-k-1}, \quad (1.2)$$

tj. když se \tilde{x} liší od x nejvýše o pět jednotek řádu příslušejícího následujícímu desetinnému místu. Pokud nerovnost (1.2) platí pro $k \leq p$, ale už neplatí pro $k = p + 1$, říkáme, že \tilde{x} má p platných desetinných míst. Správně zaokrouhlená hodnota čísla x má tedy všechna desetinná místa platná.

Z uvedené definice je zřejmé, že když je k -té desetinné místo platné, jsou platná i všechna předchozí desetinná místa.

Pro lepší pochopení pojmů platná dekadická cifra a platné desetinné místo si uvedeme několik příkladů.

x	\tilde{x}	platné cifry	platná desetinná místa
374	380	1	—
-34,5438	-34,497	3	1
100,001	99,9965	4	2
99,9965	100,001	5	2
0,873	0,871	2	2
-0,004837	-0,0053	1	3
$2,753 \cdot 10^{-8}$	$3,4 \cdot 10^{-8}$	0	7

Všimněte si, že v posledních dvou příkladech nejsou nuly za desetinnou čárkou platné cifry (není před nimi žádná nenulová cifra), ale jsou to platná desetinná místa.

Při provádění výpočtů s aproximacemi přesných hodnot může dojít ke zvětšení absolutní nebo relativní chyby výsledku a k významné ztrátě platných cifer. Jako nejkritičtější z hlediska šíření chyb se jeví odečítání dvou velmi blízkých čísel, kdy významně narůstá relativní chyba, a dělení číslem blízkým nule, kdy významně narůstá absolutní chyba. Podrobněji viz např. [13, 41, 62].

Zaokrouhlovací pravidla

Připomeňme ještě pravidla pro správné zaokrouhlování. Předpokládejme, že číslo x má (konečný nebo nekonečný) dekadický rozvoj

$$x = \pm [d_1 \cdot 10^e + d_2 \cdot 10^{e-1} + \dots + d_k \cdot 10^{e+1-k} + d_{k+1} \cdot 10^{e-k} + \dots].$$

Někdy chceme číslo x zaokrouhlit na k cifer, tj. odseknout část dekadického rozvoje začínající cifrou d_{k+1} a případně upravit poslední cifru d_k (eventuálně i předcházející, pokud je $d_k = 9$). V tom případě postupujeme následovně:

- 1) Je-li $d_{k+1} \in \{0, 1, 2, 3, 4\}$, cifru d_k neměníme.
- 2) Je-li $d_{k+1} \in \{6, 7, 8, 9\}$, cifru d_k zvětšíme o jedničku.
- 3) Je-li $d_{k+1} = 5$, rozlišíme dva případy:
 - Jestliže je některá z cifer d_{k+2}, d_{k+3}, \dots nenulová, cifru d_k zvětšíme o jedničku.
 - Jestliže jsou všechny cifry d_{k+2}, d_{k+3}, \dots nulové, cifru d_k neměníme, pokud je sudá, a zvětšíme ji o jedničku, pokud je lichá.

Číslo \tilde{x} , které takto dostaneme, bude vždy správně zaokrouhlenou hodnotou čísla x .
Uvedme několik příkladů:

počet cifer	x	\tilde{x}	počet cifer	x	\tilde{x}
2	3 245	3 200	4	0,99999	1,0000
2	3 254	3 300	3	-5,2551	-5,26
3	2,463	2,46	2	3,25	3,2
4	-31,378	-31,38	2	3,35	3,4
4	7,2396	7,240	3	-2,395	-2,40

Všimněte si, že koncové nuly v zaokrouhlených číslech 3 200 a 3 300 nejsou platné, ale musí být napsány, protože jinak by zápis nedával smysl (dostali bychom čísla 32 a 33). Naproti tomu koncové nuly v čísle 1,0000 jsou platné, a proto jsou napsány (kdybychom je vynechali, dostali bychom číslo 1 a z tohoto zápisu by nebylo možné zjistit, že číslo má pět platných cifer). Tedy u celých čísel vzniklých zaokrouhlením nelze jednoznačně říci (bez znalosti přesného čísla), kolik cifer je platných. Např. číslo 5 000 může mít jednu až čtyři platné cifry.

1.3 Reprezentace čísel v počítači

Reálná čísla jsou v počítači reprezentována v *systému čísel s pohyblivou řádovou čárkou*. V podstatě jde o semilogaritmický zápis s normalizovanou mantisou a základem q . Například $5,623 \cdot 10^5$ ($q = 10$) nebo $1,01011 \cdot 2^{-10}$ ($q = 2$). Systém těchto čísel lze charakterizovat čtyřmi celými čísly: *základem číselné soustavy* $q \geq 2$, *přesností* $p \geq 1$ (počet cifer mantisy) a *rozsahe exponentu* $[L, U]$, kde $L < 0 < U$. Označíme-li tento

systém písmenem \mathbb{F} , pak každé nenulové číslo $x \in \mathbb{F}$ má tvar

$$x = \pm m \cdot q^e, \quad \text{kde} \quad m = d_1 + \frac{d_2}{q} + \frac{d_3}{q^2} + \cdots + \frac{d_p}{q^{p-1}},$$

$d_i \in \{0, 1, \dots, q-1\}$, $i = 1, \dots, p$, $d_1 \neq 0$ a $L \leq e \leq U$. Číslo m je *normalizovaná mantisa* a e je *exponent*. Pro $x = 0$ klademe $m = e = 0$. Snadno se ověří, že množina \mathbb{F} je konečná a obsahuje $2(q-1)q^{p-1}(U-L+1) + 1$ čísel. Čísla z množiny \mathbb{F} nazýváme *strojová čísla*.

Reálné číslo x lze zapsat v počítači přesně, jen pokud $x \in \mathbb{F}$. Číslo x , které není strojové, musí být při vložení do počítače zaokrouhleno na nejbližší strojové číslo. Podobně při provádění aritmetických operací se strojovými čísly není výsledek obecně strojové číslo a musí být opět zaokrouhlen na nejbližší strojové číslo. Navíc se může stát, že vkládané číslo nebo číslo, které je výsledkem aritmetické operace, je v absolutní hodnotě příliš velké a je mimo rozsah množiny \mathbb{F} . Pak dojde k tzv. *přetečení* a výpočet je přerušen. Podrobnější informace viz [13, 22, 41, 52, 62].

V počítačích vyrobených po roce 1985 jsou čísla reprezentována podle standardu

IEEE 754-1985¹. Standard IEEE 754-2008 ze srpna 2008, který vychází z původního standardu, definuje pět základních formátů: tři binární ($q = 2$) a dva dekadické ($q = 10$). Podle počtu bitů použitých k uložení čísla v počítači, se označují binary32 (jednoduchá přesnost), binary64 (dvojnásobná přesnost), binary128 (čtyřnásobná přesnost), decimal64 a decimal128. Více podrobností lze nalézt v [71, 75].

Kromě této tzv. *hardwarové pohyblivé čárky*, kdy aritmetické výpočty jsou prováděny přímo v procesoru a jsou rychlé, se používá zápis čísel v *softwarové pohyblivé čárce*. V tomto případě jsou výpočty emulovány softwarově (pracuje se s teoreticky libovolně dlouhými řetězci cifer). Mluvíme o aritmetice s libovolnou přesností. Výpočty jsou výrazně pomalejší, ale přesnost je omezena jen množstvím volné paměti hostitelského systému. Důvodem použití softwarové pohyblivé čárky je, že hardwarová přesnost někdy není dostatečná. Viz [60].

¹Institute of Electrical and Electronics Engineers (IEEE) je profesní sdružení, jehož vedení sídlí v New Yorku a jehož posláním je podpora technologického pokroku a dokonalosti.

1.4 Korektnost a podmíněnost úlohy

Korektní úlohy

Matematickou úlohu lze obvykle chápat jako zobrazení $f(x)$, které vstupu x z množiny vstupních dat A přiřadí výstup y z množiny výstupních dat B . Řekneme, že taková úloha je *korektní*, jestliže

- 1) pro libovolná vstupní data $x \in A$ existuje jediné řešení $y = f(x) \in B$,
- 2) toto řešení spojitě závisí na vstupních datech, tedy dostatečně malé změny vstupu x vyvolávají pouze malé změny výstupu y .

Úlohy, které nejsou korektní (např. nemají pro některá vstupní data jediné řešení), nelze rozumně numericky řešit.

Podmíněnost úloh

U korektní úlohy dostatečně malá změna ve vstupních datech vyvolává malou změnu ve výstupních datech. Otázkou ovšem je, co znamenají slova „dostatečně malá“. Zda je

u konkrétní úlohy z praktického hlediska možné zajistit dostatečně malé nepřesnosti na vstupu, aby nepřesnosti vznikající díky tomu na výstupu byly přijatelné. Je tedy třeba nějakým způsobem kvantifikovat vztah mezi velikostí odchylek na vstupu a na výstupu.

Řekneme, že korektní úloha je *dobře podmíněna*, jestliže je poměr relativní chyby výstupu a vstupu malý. Předpokládejme pro jednoduchost, že vstupy A a výstupy B jsou reálná čísla. Označme:

- x vstupní údaj,
- Δx absolutní chybu na vstupu,
- $y = f(x)$ výstupní údaj,
- Δy absolutní chybu na výstupu.

Číslo

$$C_p = \left| \frac{\Delta y}{y} \right| \bigg/ \left| \frac{\Delta x}{x} \right| = \frac{|\text{relativní chyba na výstupu}|}{|\text{relativní chyba na vstupu}|}$$

nazýváme *číslem podmíněnosti* úlohy $y = f(x)$. Je-li $C_p \approx 1$, je úloha (velmi) *dobře podmíněná*, je-li $C_p \gg 1$ (např. $C_p \approx 100$), je úloha *špatně podmíněná*.

Obdobně postupujeme, jsou-li prvky množin A , B složitější, např. jsou-li to n -tice reálných čísel. Absolutní hodnoty v definici čísla podmíněnosti nahradíme normami (např. eukleidovskými, které jsou indukované standardním skalárním součinem v \mathbb{R}^n , viz definice 1.3).

Výpočet čísla podmíněnosti si ukážeme na příkladu.

Příklad 1.1 Uvažujme soustavu lineárních rovnic o dvou neznámých a soustavu, která z ní vznikla malou změnou jednoho koeficientu:

$$\begin{array}{rcl} a + b & = & 2, \\ a + 1,01b & = & 2,01, \end{array} \qquad \begin{array}{rcl} a + b & = & 2, \\ a + 1,01b & = & 2,02. \end{array}$$

Posuďte podmíněnost výchozí soustavy lineárních rovnic.

Řešení. Nejprve ověříme, že úloha je korektní. Obě soustavy lze zapsat jedním zápisem

$$\begin{array}{rcl} a + b & = & 2, \\ a + 1,01b & = & x, \end{array}$$

kde za x dosadíme 2,01 nebo 2,02. Vstupem je tudíž parametr x a výstupem dvojice řešení (a, b) .

Snadno se ověří, že soustava má pro libovolné x jediné řešení

$$a = \frac{2,02 - x}{0,01}, \qquad b = \frac{x - 2}{0,01}.$$

Z předchozích vzorců je navíc vidět, že výstup (a, b) spojitě závisí na vstupu x .

Pro uvažované dvě hodnoty parametru x dostaneme:

$$x = 2,01 \quad \Rightarrow \quad a = 1, b = 1, \quad x = 2,02 \quad \Rightarrow \quad a = 0, b = 2.$$

Tedy malá změna druhého koeficientu pravé strany vedla k podstatně jinému řešení soustavy.

Určíme číslo podmíněnosti:

$$\text{Vstup: } x = 2,01, |x| = 2,01, \Delta x = 2,01 - 2,02 = -0,01, |\Delta x| = 0,01,$$

$$\begin{aligned} \text{Výstup: } y &= (1, 1), \|y\| = \sqrt{1^2 + 1^2} = \sqrt{2}, \Delta y = (1, 1) - (0, 2) = (1, -1), \\ \| \Delta y \| &= \sqrt{1^2 + (-1)^2} = \sqrt{2}, \end{aligned}$$

takže

$$C_p = \frac{\| \Delta y \|}{\| y \|} \bigg/ \frac{|\Delta x|}{|x|} = \frac{\sqrt{2}}{\sqrt{2}} \bigg/ \frac{0,01}{2,01} = \frac{2,01}{0,01} = 201.$$

Úloha je tedy špatně podmíněná.

Připomeňme ještě Cramerovo pravidlo pro řešení čtvercových soustav lineárních rovnic s regulární maticí soustavy. Vzorce pro neznámé mají tvar zlomku, kde ve jmenovateli je determinant matice soustavy. Ten je v našem případě $\begin{vmatrix} 1 & 1 \\ 1 & 1,01 \end{vmatrix} = 0,01$ (srovnejte vzorce pro a a b), což je malé číslo. To je příčinou špatné podmíněnosti a numerických problémů při řešení naší soustavy. ▲

1.5 Vlastnosti numerických algoritmů

Jak již bylo řečeno v úvodu, postup, který vede k řešení numerické úlohy, se nazývá numerický algoritmus. Co je obecně algoritmus (přesněji co jsme ochotni považovat za algoritmus), není jednoduchá otázka (viz *Churchova¹-Turingova² teze*, [70]) a není

¹**Alonzo Church** (1903–1995) (čti čerč) — americký matematik a logik, který přispěl významně k rozvoji matematické logiky a teoretických základů informatiky.

²**Alan Mathison Turing** (1912–1954) (čti turing) — britský matematik, logik a expert na dešifrování, který významně přispěl k rozvoji informatiky. Během 2. světové války se rozhodujícím způsobem podílel na rozluštění kódu německého šifrovacího stroje Enigma.

naším úkolem ji řešit. Zabývá se jí teorie algoritmů, která je součástí algebry a teoretické informatiky. Intuitivně za algoritmus považujeme jakýsi postup, který slouží k řešení nějakého problému. Příkladů algoritmů z matematiky i jiných oblastí známe řadu, např. ruční postupy pro sečítání, odčítání, násobení a dělení, Eukleidův algoritmus pro nalezení společného dělitele, Hornerovo schéma, dělení polynomů se zbytkem, návod na instalaci softwaru, návod jak vymalovat pokoj, návod jak pěstovat květinu atd.

Nás budou zajímat *numerické algoritmy*, tedy algoritmy, jejichž vstupy i výstupy jsou konečné množiny čísel a kde kroky algoritmů se provádějí na počítači. Při tom, jak již víme, vznikají zaokrouhlovací chyby, nejprve při ukládání vstupních hodnot a pak při provádění aritmetických operací. Aby výstup nebyl zcela znehodnocen a získané výsledky nebyly nesmyslné, je třeba, aby algoritmus byl tzv. *stabilní*. To znamená, že musí být

- 1) *dobře podmíněný*, tj. málo citlivý na poruchy ve vstupních datech,
- 2) *numericky stabilní*, tj. málo citlivý na vliv zaokrouhlovacích chyb vznikajících během výpočtu.

1.6 Symbol O

V numerické matematice (i v jiných oblastech matematiky, např. teorii čísel, teorii složitosti algoritmů, v analýze a obecně při studiu asymptotických vlastností) je často potřebné vyjádřit, že jedna veličina se zmenšuje rychleji nebo roste pomaleji než jiná veličina. K tomu se používá speciální symbol O , často označovaný jako *Landauův¹ symbol velké O* . Formální definice je následující:

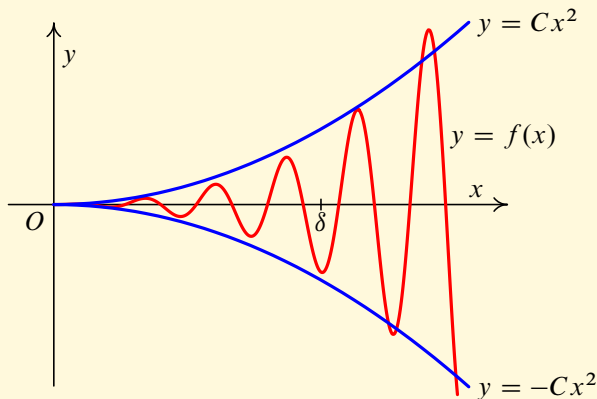
Definice 1.2 Nechtě $f(x)$ a $g(x)$ jsou funkce definované v okolí bodu x_0 (připouštíme i $x_0 = \pm\infty$). Řekneme, že $f(x) = O(g(x))$ v bodě x_0 (čteme f je velké O g), jestliže existuje konstanta $C > 0$ taková, že $|f(x)| \leq C|g(x)|$ v nějakém (jednostranném nebo oboustranném) okolí bodu x_0 .

Tedy blízko bodu x_0 nemůže být $|f(x)|$ v jistém smyslu větší než $|g(x)|$. Říkáme rovněž, že *funkce $f(x)$ je v bodě x_0 nejvýše stejného řádu jako funkce $g(x)$* .

Nás bude nejvíce zajímat případ, kdy $x_0 = 0$ a funkce $g(x)$ bude mocnina proměnné x .

¹**Edmund Georg Hermann Landau** (1877–1938) — německý matematik, který se zabýval teorií čísel a funkcemi komplexní proměnné.

Např. zápis $f(x) = O(x^2)$ v bodě nula znamená, že existuje konstanta $C > 0$ taková, že v dostatečně malém okolí nuly platí $|f(x)| \leq Cx^2$, tj. $-Cx^2 \leq f(x) \leq Cx^2$. Situace je znázorněna na obr. 1.1; na intervalu $\langle 0, \delta \rangle$ je graf funkce $f(x)$ mezi grafy funkcí $-Cx^2$ a Cx^2 .



Obr. 1.1: Význam symbolu $f(x) = O(g(x))$ pro $g(x) = x^2$

Při počítání se symbolem O musíme být obezřetní. Např. ze vztahů $f_1(x) = O(g(x))$ a $f_2(x) = O(g(x))$ v bodě x_0 rozhodně nelze usuzovat, že platí $f_1(x) = f_2(x)$. Zmíněné zápisy totiž pouze značí, že existují kladné konstanty C_1 a C_2 takové, že $|f_1(x)| \leq C_1|g(x)|$ a $|f_2(x)| \leq C_2|g(x)|$ v nějakém okolí bodu x_0 . Např. v bodě nula platí $x \sin x = O(x)$ a $x \cos x = O(x)$, ale určitě není pravda, že $x \sin x = x \cos x$.

Podobně musíme chápat např. zápisy v bodě nula

$$O(x^k) \pm O(x^k) = O(x^k), \quad x^l O(x^k) = O(x^{k+l}), \quad O(x^k)O(x^l) = O(x^{k+l})$$

apod. Např. první z nich znamená, že pokud $f(x) = O(x^k)$ a $g(x) = O(x^k)$, pak rovněž platí $f(x) \pm g(x) = O(x^k)$.

Protože jiný případ než $x_0 = 0$ a pouze pravé okolí nebudeme až na výjimky potřebovat, budeme slova „v bodě nula“ obvykle vynechávat. Existují i další Landauovy symboly, o (malé o), Ω , ω , Θ a \sim , vyjadřující různé vztahy mezi rychlostmi změn dvou veličin — viz [61]. Symbol o je používán často v analýze, viz [33, str. 301].

1.7 Vektorové a maticové normy, skalární součin

Chceme-li popsat, jak moc se liší dvě reálná čísla x a y , řekneme, jaká je jejich vzdálenost. Ta je, jak známo, dána absolutní hodnotou jejich rozdílu $|x - y|$. Jsou-li tedy tato čísla blízka, je $|x - y|$ malé číslo. Jestliže nekonečná posloupnost čísel $\{x_k\}$ má limitu x , znamená to, že $\lim_{k \rightarrow \infty} |x_k - x| = 0$, tedy že vzdálenosti čísel x_k od čísla x se s rostoucím k přibližují k nule.

Připomeňme, jak je absolutní hodnota definovaná a jaké má vlastnosti. Pro reálné číslo x klademe

$$|x| = \begin{cases} x & \text{pro } x \geq 0, \\ -x & \text{pro } x < 0. \end{cases}$$

Hodnota $|x|$ udává vzdálenost čísla x od nuly, tj. od počátku souřadnic na číselné ose. Zřejmě $|x| \geq 0$, přičemž $|x| = 0$ právě tehdy, když $x = 0$.

Pro libovolná dvě reálná čísla x, y platí následující vztahy:

$$\begin{aligned}
 |-x| &= |x|, & |x \pm y| &\leq |x| + |y|, & |x \pm y| &\geq ||x| - |y||, \\
 |xy| &= |x| \cdot |y|, & \left| \frac{x}{y} \right| &= \frac{|x|}{|y|} & (y \neq 0).
 \end{aligned}$$

Hodnota $|x - y|$ udává, jak již bylo řečeno, vzdálenost čísel x a y na číselné ose.

1.7.1 Vektorové normy

V dalším textu budeme pracovat s uspořádanými n -ticemi reálných čísel x_1, x_2, \dots, x_n , kde $n \in \mathbb{N}$. Ty lze chápat jako řádkové matice $[x_1, \dots, x_n]$ nebo sloupcové matice $[x_1, \dots, x_n]^T$ (symbol T značí transponování). Často se pro ně rovněž používá název (aritmetické) vektory. Rádi bychom také dovedli vyjádřit, že jejich vzdálenost je malá, že se posloupnost takových n -tic přibližuje k nějaké n -tici apod. Pro $n = 2$ a $n = 3$ můžeme dvojice resp. trojice reálných čísel chápat jako souřadnice bodů v rovině resp. prostoru. V těchto případech máme názornou představu, co znamená jejich vzdálenost. Pro $n \geq 4$ tomu už tak není.

Abychom mohli nějakým způsobem měřit velikost a vzdálenost takových vektorů, zavedeme následující důležitý pojem, jehož definice je motivovaná některými vlastnostmi absolutní hodnoty. Připomeňme, že pro každé dvě n -tice \mathbf{x} , \mathbf{y} (chápané jako matice) je definován jejich *součet* $\mathbf{x} + \mathbf{y}$ (sčítají se složky vektorů se stejnými indexy) a pro každé číslo c a každou n -tici \mathbf{x} je definován (skalární) násobek vektoru číslem $c\mathbf{x}$ (číslem c se násobí každá složka vektoru).

V dalším textu bude vhodnější, když budeme vektory považovat za sloupcové matice. Tedy n -rozměrný vektor označíme $\mathbf{x} = [x_1, \dots, x_n]^T$ (zápis pomocí transponování je vhodnější, zabere v textu méně místa). Speciálně $\mathbf{0}$ bude nulový vektor, tj. $\mathbf{0} = [0, \dots, 0]^T$. Množinu všech takových sloupcových vektorů označíme \mathbb{R}^n .

Definice 1.3 Funkce, která každému vektoru $\mathbf{x} \in \mathbb{R}^n$ přiřazuje reálné číslo $\|\mathbf{x}\|$, se nazývá (vektorová) *norma na \mathbb{R}^n* , jestliže má následující vlastnosti:

- 1) $\|\mathbf{x}\| \geq 0$ pro každý vektor $\mathbf{x} \in \mathbb{R}^n$, přičemž $\|\mathbf{x}\| = 0$, právě když $\mathbf{x} = \mathbf{0}$;
- 2) $\|c\mathbf{x}\| = |c| \cdot \|\mathbf{x}\|$ pro každé číslo $c \in \mathbb{R}$ a každý vektor $\mathbf{x} \in \mathbb{R}^n$;
- 3) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ pro každé vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Druhá vlastnost se (někdy) nazývá *homogenita*, třetí pak *trojúhelníková nerovnost* nebo *subaditivita*.

Na množině \mathbb{R}^n existuje nekonečně mnoho různých norem. Pokud budeme pracovat s více normami najednou, označíme je vhodným dolním indexem, např. $\|\cdot\|_a$ a $\|\cdot\|_b$, kde a, b jsou nějaké symboly. Uvedeme nyní příklady nejdůležitějších norem, které se praxi používají.

Nechť $p \geq 1$ je reálné číslo. Pak funkce definovaná vztahem

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \quad (1.3)$$

je vektorová norma. Ověření prvních dvou vlastností normy je snadné, důkaz trojúhelníkové nerovnosti viz např. [15, str. 70].

Nejčastěji užívané speciální případy jsou

$$\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n| \quad (\text{součtová norma}),$$

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2} \quad (\text{eukleidovská norma}),$$

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, \dots, |x_n|\} \quad (\text{maximální norma}).$$

Třetí norma není speciálním případem vzorce (1.3), ale snadno se ověří pomocí věty o limitě tří funkcí (viz [16, str. 70] nebo [33, str. 156]), že $\lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty$. Označení je tedy logické.

Příklad 1.4 Je dán vektor $\mathbf{x} = [-1, 2, 3, -2]^T \in \mathbb{R}^4$. Vypočtěte jeho normy $\|\mathbf{x}\|_p$ pro $p = 1, 2, \infty$.

Řešení. Platí:

$$\|\mathbf{x}\|_1 = |-1| + |2| + |3| + |-2| = 8,$$

$$\|\mathbf{x}\|_2 = \sqrt{(-1)^2 + 2^2 + 3^2 + (-2)^2} = \sqrt{18},$$

$$\|\mathbf{x}\|_\infty = \max\{|-1|, |2|, |3|, |-2|\} = 3. \quad \blacktriangle$$

Pro dva vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ se číslo $\|\mathbf{x} - \mathbf{y}\|$ nazývá *vzdálenost vektorů \mathbf{x}, \mathbf{y} v normě $\|\cdot\|$* (srovnejte se vzorcem pro vzdálenost dvou čísel). Dále pro vektor $\mathbf{x}_0 \in \mathbb{R}^n$ a číslo $r > 0$ se množina $K(\mathbf{x}_0, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| \leq r\}$ nazývá *uzavřená koule v normě $\|\cdot\|$ se středem \mathbf{x}_0 a poloměrem r* . Je to tedy množina všech vektorů, které mají od pevného vektoru \mathbf{x}_0 vzdálenost nejvýše r . Podobně definujeme *otevřenou kouli*

$O(\mathbf{x}_0, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| < r\}$ a kulovou plochu $S(\mathbf{x}_0, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| = r\}$. Je-li $n = 2$, používáme místo názvu koule název kruh a místo názvu kulová plocha název kružnice. Pojmenování koule a kulová plocha resp. kruh a kružnice musíme brát s rezervou. Vzdálenost se měří pomocí dané normy, takže tvar těchto množin obecně neodpovídá tomu, co standardně tato slova označují, viz následující příklad.

Příklad 1.5 Nakreslete obrázky kruhů se středem v počátku a poloměrem $r > 0$ v \mathbb{R}^2 v normách $\|\cdot\|_p$ pro $p = 1, 2, \infty$.

Řešení. Prvky množiny \mathbb{R}^2 lze ztotožnit s kartézskými souřadnicemi bodů v rovině. Hledáme tedy dvojice $\mathbf{x} = [x_1, x_2]^T$, pro něž platí $\|\mathbf{x} - \mathbf{0}\| = \|\mathbf{x}\| \leq r$ v uvedených normách.

1) Pro $p = 1$ dostaneme nerovnost $\|\mathbf{x}\|_1 = |x_1| + |x_2| \leq r$. Zjistíme, jak tato podmínka vypadá v jednotlivých kvadrantech.

V prvním kvadrantu je $x_1 \geq 0$ a $x_2 \geq 0$, takže má platit $x_1 + x_2 \leq r$. Body tudíž leží pod přímkou o rovnici $x_2 = -x_1 + r$.

Ve druhém kvadrantu je $x_1 \leq 0$ a $x_2 \geq 0$, takže má platit $-x_1 + x_2 \leq r$. Body tudíž leží pod přímkou o rovnici $x_2 = x_1 + r$.

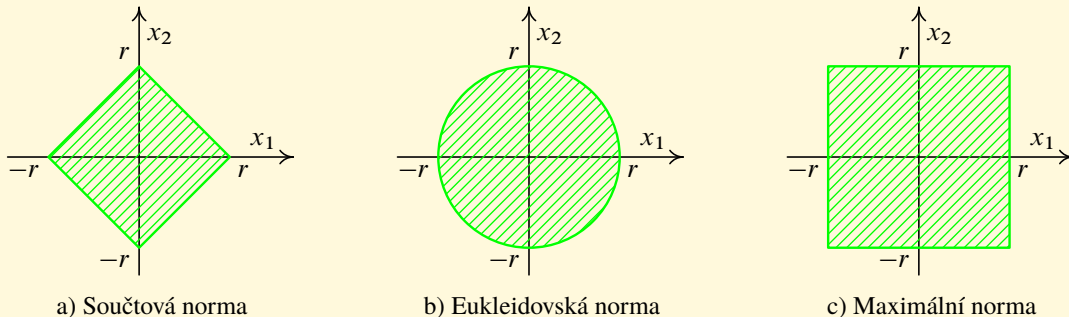
Ve třetím kvadrantu je $x_1 \leq 0$ a $x_2 \leq 0$, takže má platit $-x_1 - x_2 \leq r$. Body tudíž leží nad přímkou o rovnici $x_2 = -x_1 - r$.

Ve čtvrtém kvadrantu je $x_1 \geq 0$ a $x_2 \leq 0$, takže má platit $x_1 - x_2 \leq r$. Body tudíž leží nad přímkou o rovnici $x_2 = x_1 - r$.

Výsledek je znázorněn na obr. 1.2 a). Jedná se o čtverec s vrcholy $[r, 0]$, $[0, r]$, $[-r, 0]$ a $[0, -r]$.

2) Pro $p = 2$ dostaneme nerovnost $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2} \leq r$, tj. $x_1^2 + x_2^2 \leq r^2$. Jedná se tedy o kruh v normálním slova smyslu, viz obr. 1.2 b), což se dalo čekat, protože jde o eukleidovskou normu.

3) Pro $p = \infty$ dostaneme nerovnost $\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|\} \leq r$. Musí tudíž platit $|x_1| \leq r$ i $|x_2| \leq r$, tj. $-r \leq x_1 \leq r$ a současně $-r \leq x_2 \leq r$. Body splňující první podmínku leží mezi svislými přímkami o rovnicích $x_1 = -r$ a $x_1 = r$ a body splňující druhou podmínku leží mezi vodorovnými přímkami o rovnicích $x_2 = -r$ a $x_2 = r$. Výsledek je znázorněn na obr. 1.2 c). Jedná se o čtverec s vrcholy $[r, r]$, $[-r, r]$, $[-r, -r]$ a $[r, -r]$. ▲

Obr. 1.2: Kruh v \mathbb{R}^2 v různých normách

V dalším textu budeme pracovat s posloupnostmi vektorů. Protože dolními indexy jsou očíslovány složky vektorů, použijeme, tak jak je obvyklé, pro označení pořadí členů posloupnosti horní indexy v kulatých závorkách. Tedy

$$\mathbf{x}^{(k)} = [x_1^{(k)}, \dots, x_n^{(k)}]^T, \quad \text{kde } k \in \mathbb{N}.$$

Definice 1.6 Nechť $\{\mathbf{x}^{(k)}\}$, $k = 1, 2, \dots$, je posloupnost vektorů v \mathbb{R}^n . Řekneme, že tato posloupnost *konverguje v normě* $\|\cdot\|$ k vektoru $\mathbf{x} \in \mathbb{R}^n$ a píšeme $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{x}$ nebo $\mathbf{x}^{(k)} \rightarrow \mathbf{x}$ pro $k \rightarrow \infty$, právě když $\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}\| = 0$.

Konvergence v normě je tedy definována pomocí konvergence číselné posloupnosti $\|\mathbf{x}^{(k)} - \mathbf{x}\|$. Její hodnoty však závisí na volbě normy. Je tedy přirozené položit si otázku, zda je možné, že nějaká posloupnost v jedné normě konverguje, ale v jiné ne. K odpovědi na tuto otázku zavedeme následující pojem.

Definice 1.7 Řekneme, že dvě normy $\|\cdot\|_a$ a $\|\cdot\|_b$ na \mathbb{R}^n jsou *ekvivalentní*, jestliže existují kladné konstanty c_1 a c_2 takové, že pro libovolný vektor $\mathbf{x} \in \mathbb{R}^n$ platí nerovnosti $c_1 \|\mathbf{x}\|_a \leq \|\mathbf{x}\|_b \leq c_2 \|\mathbf{x}\|_a$.

Je zřejmé, že pro ekvivalentní normy platí, že $\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}\|_a = 0$, právě když $\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}\|_b = 0$. Tedy ekvivalentní normy určují tytéž konvergentní posloupnosti

vektorů. Jako příklad ekvivalentních norem lze uvést dvojici $\|\cdot\|_1$ a $\|\cdot\|_\infty$. Platí totiž:

$$\begin{aligned}\|\mathbf{x}\|_\infty &= \max\{|x_1|, \dots, |x_n|\} = |x_r| \leq |x_1| + \dots + |x_n| = \\ &= \|\mathbf{x}\|_1 \leq \underbrace{|x_r| + \dots + |x_r|}_{n \text{ krát}} = n|x_r| = n\|\mathbf{x}\|_\infty.\end{aligned}$$

Lze tedy zvolit $c_1 = 1$ a $c_2 = n$. Tyto konstanty nelze zlepšit (c_1 zvětšit a c_2 zmenšit), protože pro $\mathbf{x} = [1, 0, \dots, 0]^T$ platí $\|\mathbf{x}\|_\infty = 1 = \|\mathbf{x}\|_1$ a pro $\mathbf{x} = [1, 1, \dots, 1]^T$ platí $\|\mathbf{x}\|_\infty = 1$ a $\|\mathbf{x}\|_1 = n$, tedy $\|\mathbf{x}\|_1 = n\|\mathbf{x}\|_\infty$.

Lze dokázat následující netriviální výsledek (viz např. [8, str. 157], [52, str. 209] nebo [54, str. 93 a 100]).

Věta 1.8 *Libovolné dvě vektorové normy na \mathbb{R}^n jsou ekvivalentní.*

Z předchozí věty vyplývá, že pokud nějaká posloupnost vektorů konverguje v jedné normě, konverguje v jakékoli jiné a limita je vždy stejná. Dále je snadno vidět, že posloupnost vektorů $\mathbf{x}^{(k)}$ konverguje v normě $\|\cdot\|_1$ k vektoru \mathbf{x} , právě když posloupnosti jednotlivých složek $x_i^{(k)}$, $i = 1, \dots, n$, konvergují ke složkám x_i . Ve spojení s větou 1.8 dostáváme následující důležitý výsledek.

Důsledek 1.9 Posloupnost vektorů $\mathbf{x}^{(k)}$ konverguje v nějaké vektorové normě na \mathbb{R}^n k vektoru \mathbf{x} , právě když $\lim_{k \rightarrow \infty} x_i^{(k)} = x_i$ pro každé $i = 1, \dots, n$.

Tedy konvergence posloupnosti vektorů v libovolné normě je ekvivalentní s konvergencí číselných posloupností složek (mluvíme o konvergenci po složkách). Zdánilivě by tudíž stačilo vybrat a používat jen jednu normu. Ukazuje se však, že to není pravda, v různých situacích je výhodné pracovat s různými normami.

Poznámka 1.10 Doposud jsme pracovali s vektory, jejichž složky byla reálná čísla. Analogicky je možné uvažovat vektory, jejichž složky jsou komplexní čísla. Množinu všech takových sloupcových matic označíme \mathbb{C}^n . Ukazuje se, že všechny pojmy, které jsme zavedli v tomto oddílu, je možné beze změn přenést i na komplexní případ. (U eukleidovské normy musí být místo x_i^2 , což nemusí být reálné číslo, $|x_i|^2$.) Rovněž všechna uvedená tvrzení zůstávají v platnosti.

1.7.2 Maticové normy

Podobně jako je v nejrůznějších částech matematiky třeba pracovat s řádky nebo sloupci čísel, je často potřebné pracovat se soubory čísel uspořádanými do obdélníkových schémat,

tedy s maticemi. I v případě matic je žádoucí mít možnost posuzovat, jak je daná matice „velká“ nebo jak jsou dvě matice „vzdálené“.

Označme $\mathbb{M}_{m,n}$ množinu všech obdélníkových matic o m řádcích a n sloupcích s reálnými prvky. Je-li $m = n$, tj. jde-li o čtvercové matice, použijeme označení \mathbb{M}_n . Připomeňme, že pro libovolné matice $A, B \in \mathbb{M}_{m,n}$ a libovolné číslo $c \in \mathbb{R}$ je definován součet matic $A + B$ a násobek matice číslem cA .

Naskládáme-li prvky matice z množiny $\mathbb{M}_{m,n}$ do jednoho sloupce, dostaneme vlastně vektor z množiny $\mathbb{R}^{m \times n}$. Při tomto ztotožnění si budou odpovídat i operace sečítání matic a násobení matice číslem. Nepřekvapí proto, že definice normy matice bude téměř identická s definicí normy vektoru (což je vlastně sloupcová matice).

Definice 1.11 Funkce, která každé matici $A \in \mathbb{M}_{m,n}$ přiřazuje reálné číslo $\|A\|$, se nazývá (maticová) *norma na $\mathbb{M}_{m,n}$* , jestliže má následující vlastnosti:

- 1) $\|A\| \geq 0$ pro každou matici $A \in \mathbb{M}_{m,n}$, přičemž $\|A\| = 0$ právě tehdy, když $A = O$;
- 2) $\|cA\| = |c| \cdot \|A\|$ pro každé číslo $c \in \mathbb{R}$ a každou matici $A \in \mathbb{M}_{m,n}$;
- 3) $\|A + B\| \leq \|A\| + \|B\|$ pro každé matice $A, B \in \mathbb{M}_{m,n}$.

Druhá vlastnost se (někdy) nazývá *homogenita*, třetí pak *trojúhelníková nerovnost* nebo *subaditivita*.

Na množině $\mathbb{M}_{m,n}$ existuje nekonečně mnoho různých norem. Pokud budeme pracovat s více normami najednou, označíme je vhodným dolním indexem, např. $\|\cdot\|_a$ a $\|\cdot\|_b$, kde a, b jsou nějaké symboly. Uvedeme nyní příklady tří norem, které jsou analogiemi vektorových norem $\|\cdot\|_1$, $\|\cdot\|_2$ a $\|\cdot\|_\infty$.

$$\|A\|_S = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}| \quad (\text{součtová norma}),$$

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} \quad (\text{Frobeniova}^1 \text{ norma}),$$

$$\|A\|_M = \max_{\substack{i=1,\dots,m, \\ j=1,\dots,n}} \{|a_{ij}|\} \quad (\text{maximální norma}).$$

¹**Ferdinand Georg Frobenius** (1849–1917) — německý matematik. Zabýval se zejména algebrou.

Frobeniova norma je také známá pod názvy Hilbertova¹-Schmidtova² norma nebo Schurova³ norma.

Pro maticové normy se zavádějí obdobné pojmy a platí obdobné výsledky jako pro vektorové normy. Nebudeme je podrobně vypisovat, uvedeme jen stručný přehled.

1. Pro posloupnost matic $A^{(k)} \in \mathbb{M}_{m,n}$, $k = 1, 2, \dots$, a matici $A \in \mathbb{M}_{m,n}$ definujeme, že $\lim_{k \rightarrow \infty} A^{(k)} = A$ v normě $\|\cdot\|$, právě když $\lim_{k \rightarrow \infty} \|A^{(k)} - A\| = 0$ (definice 1.6).
2. Stejným způsobem se zavádí ekvivalentní maticové normy (definice 1.7). Ty určují tytéž konvergentní posloupnosti matic v $\mathbb{M}_{m,n}$.
3. Všechny maticové normy v $\mathbb{M}_{m,n}$ jsou ekvivalentní (věta 1.8).

¹**David Hilbert** (1862–1943) — významný německý matematik. Ovlivnil řadu matematických oborů. Je považován za jednoho z nejvýznamnějších matematiků všech dob. Na 2. mezinárodním kongresu matematiků v Paříži v roce 1900 předložil 23 otevřených problémů, dnes nazývaných **Hilbertovy problémy**, které považoval za klíčové pro další rozvoj matematiky. Ne všechny byly dodnes vyřešeny.

²**Erhard Schmidt** (1876–1959) (čti šmid) — německý matematik. Zabýval se integrálními rovnicemi a topologií.

³**Issai Schur** (1875–1941) (čti šur) — německý matematik židovského původu narozený v Rusku. Zabýval se reprezentací grup.

4. Posloupnost matic $A^{(k)} \in \mathbb{M}_{m,n}$ konverguje v nějaké normě k matici $A \in \mathbb{M}_{m,n}$ právě tehdy, když konvergují posloupnosti jednotlivých složek, tj. $\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij}$, $i = 1, \dots, m$, $j = 1, \dots, n$ (důsledek 1.9).

Pro matice však existuje ještě jedna důležitá operace, a to násobení matic. Pro matici $A \in \mathbb{M}_{m,n}$ a vektor $\mathbf{x} \in \mathbb{R}^n$ je jejich součin $A\mathbf{x}$ vektorem v \mathbb{R}^m a pro matice $A \in \mathbb{M}_{m,n}$ a $B \in \mathbb{M}_{n,p}$ je jejich součin AB maticí v $\mathbb{M}_{m,p}$. Na začátku oddílu 1.7 jsme mluvili o absolutní hodnotě, která byla inspirací pro definici normy. Jednou z vlastností absolutní hodnoty je, že pro součin dvou čísel platí $|ab| = |a| \cdot |b|$. Tento požadavek však pro normy nelze rozumně splnit. Nicméně v mnoha aplikacích je podstatné umět odhadnout normu součinu pomocí norem činitelů. Proto se zavádějí následující dva důležité pojmy.

Definice 1.12

1) Nechť $\|\cdot\|_a$ je vektorová norma na \mathbb{R}^n , $\|\cdot\|_b$ je vektorová norma na \mathbb{R}^m a $\|\cdot\|$ je maticová norma na $\mathbb{M}_{m,n}$.

Řekneme, že *maticová norma* $\|\cdot\|$ je *souhlasná s vektorovými normami* $\|\cdot\|_a$ a $\|\cdot\|_b$, jestliže nerovnost $\|A\mathbf{x}\|_b \leq \|A\| \cdot \|\mathbf{x}\|_a$ platí pro libovolnou matici $A \in \mathbb{M}_{m,n}$ a libovolný vektor $\mathbf{x} \in \mathbb{R}^n$.

Speciálně, je-li $\|\cdot\|_a = \|\cdot\|_b$ (nemusí nutně platit $m = n$, touto rovností se myslí, že normy jsou dány obdobnými vzorci, viz např. (1.3)), říkáme, že *maticová norma* $\|\cdot\|$ je *souhlasná s vektorovou normou* $\|\cdot\|_a$.

2) Nechť $\|\cdot\|$ je maticová norma definovaná na $\mathbb{M}_{m,n}$ pro libovolné rozměry m a n .

Řekneme, že *maticová norma* $\|\cdot\|$ je *submultiplikativní*, jestliže nerovnost $\|AB\| \leq \|A\| \cdot \|B\|$ platí pro libovolnou dvojici matic $A \in \mathbb{M}_{m,n}$ a $B \in \mathbb{M}_{n,p}$.

Požadavek submultiplikativity se často zahrnuje do definice 1.11 maticové normy.

Lze např. ukázat, že Frobeniova maticová norma je souhlasná s eukleidovskou vektorovou normou, tj. platí $\|A\mathbf{x}\|_2 \leq \|A\|_F \cdot \|\mathbf{x}\|_2$. Dále platí, že Frobeniova norma je submultiplikativní, tedy $\|AB\|_F \leq \|A\|_F \cdot \|B\|_F$.

Konstrukce operátorové normy

V obecném případě však bohužel výše uvedené dvě důležité vlastnosti norem neplatí. Proto si nyní popíšeme konstrukci, která ke dvojici vektorových norem přiřadí jistou „nejmenší“ souhlasnou maticovou normu.

Nechť $\|\cdot\|_a$ je vektorová norma na \mathbb{R}^n , $\|\cdot\|_b$ je vektorová norma na \mathbb{R}^m a $A \in \mathbb{M}_{m,n}$ je libovolná matice. Uvažujme množinu všech jednotkových vektorů v \mathbb{R}^n , tj. takových vektorů $\mathbf{x} \in \mathbb{R}^n$, pro něž platí $\|\mathbf{x}\|_a = 1$. Tyto vektory tvoří jednotkovou kulovou plochu $S(\mathbf{0}, 1)$ v normě $\|\cdot\|_a$ v \mathbb{R}^n . Lze ukázat, že reálná nezáporná funkce $\|A\mathbf{x}\|_b$ na $S(\mathbf{0}, 1)$ nabývá největší i nejmenší hodnotu. Největší hodnotu označíme symbolem $\|A\|_{a,b}$. Tedy

$$\|A\|_{a,b} = \max_{\|\mathbf{x}\|_a=1} \|A\mathbf{x}\|_b. \quad (1.4)$$

Věta 1.13 *Funkce $\|A\|_{a,b}$ definovaná vztahem (1.4) je maticová norma na $\mathbb{M}_{m,n}$.*

Důkaz. Zřejmě $\|A\|_{a,b} \geq 0$ a $\|O\|_{a,b} = 0$. Je-li $A \neq O$, existuje $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_a = 1$, takový, že $A\mathbf{x} \neq \mathbf{0}$, takže $\|A\mathbf{x}\|_b > 0$. Tedy musí platit $\|A\|_{a,b} > 0$.

Pro $c \in \mathbb{R}$ je $\|cA\mathbf{x}\|_b = |c| \cdot \|A\mathbf{x}\|_b$, takže $\max_{\|\mathbf{x}\|_a=1} \|cA\mathbf{x}\|_b = |c| \max_{\|\mathbf{x}\|_a=1} \|A\mathbf{x}\|_b$. To znamená, že $\|cA\|_{a,b} = |c| \cdot \|A\|_{a,b}$.

Konečně pro libovolný vektor $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_a = 1$, je $\|(A + B)\mathbf{x}\|_b = \|A\mathbf{x} + B\mathbf{x}\|_b \leq \|A\mathbf{x}\|_b + \|B\mathbf{x}\|_b \leq \|A\|_{a,b} + \|B\|_{a,b}$, takže musí platit nerovnost $\|A + B\|_{a,b} \leq \|A\|_{a,b} + \|B\|_{a,b}$. \square

O maticové normě $\|\cdot\|_{a,b}$ říkáme, že je *indukovaná vektorovými normami* $\|\cdot\|_a$ a $\|\cdot\|_b$ nebo že je *přidružená k vektorovým normám* $\|\cdot\|_a$ a $\|\cdot\|_b$.

Je-li $\|\cdot\|_a = \|\cdot\|_b$ (nemusí platit $m = n$, touto rovností se myslí, že normy jsou dány obdobnými vzorci), používáme (pokud nemůže dojít nedorozumění) pro normu matice A rovněž označení $\|A\|_a$ a říkáme, že *maticová norma* $\|\cdot\|_a$ je *indukovaná vektorovou normou* $\|\cdot\|_a$ nebo že *maticová norma* $\|\cdot\|_a$ je *přidružená k vektorové normě* $\|\cdot\|_a$.

Maticová norma indukovaná nějakými vektorovými normami se nazývá *operátorová maticová norma*.

Uvedeme dva další vztahy, pomocí nichž lze ekvivalentně maticovou normu $\|\cdot\|_{a,b}$ definovat.

Pro nenulový vektor \mathbf{x} platí $\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_a} \right\|_a = \left\| \frac{1}{\|\mathbf{x}\|_a} \mathbf{x} \right\|_a = \frac{1}{\|\mathbf{x}\|_a} \|\mathbf{x}\|_a = 1$ (použili jsme homogenitu normy). Dále $A\left(\frac{\mathbf{x}}{\|\mathbf{x}\|_a}\right) = A\left(\frac{1}{\|\mathbf{x}\|_a} \mathbf{x}\right) = \frac{1}{\|\mathbf{x}\|_a} A\mathbf{x}$ (podle pravidel pro násobení matice číslem). Tedy (opět použijeme homogenitu normy) platí, že $\left\| A\left(\frac{\mathbf{x}}{\|\mathbf{x}\|_a}\right) \right\|_b =$

$= \left\| \frac{1}{\|\mathbf{x}\|_a} A\mathbf{x} \right\|_b = \frac{1}{\|\mathbf{x}\|_a} \|A\mathbf{x}\|_b$. Probíhá-li \mathbf{x} všechny nenulové vektory v \mathbb{R}^n , pak $\frac{\mathbf{x}}{\|\mathbf{x}\|_a}$ probíhá všechny vektory v $S(\mathbf{0}, 1)$. Platí tudíž, že

$$\|A\|_{a,b} = \max_{\|\mathbf{x}\|_a=1} \|A\mathbf{x}\|_b = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_b}{\|\mathbf{x}\|_a}. \quad (1.5)$$

Z předchozího vztahu plyne důležitý poznatek, že pro libovolný nenulový vektor $\mathbf{x} \in \mathbb{R}^n$ je splněna nerovnost $\|A\mathbf{x}\|_b \leq \|A\|_{a,b} \cdot \|\mathbf{x}\|_a$. Ta platí triviálně i pro nulový vektor. Přitom číslo $\|A\|_{a,b}$ je nejmenší konstanta k , pro niž platí pro všechna $\mathbf{x} \in \mathbb{R}^n$ nerovnost $\|A\mathbf{x}\|_b \leq k\|\mathbf{x}\|_a$. Dostáváme následující tvrzení.

Důsledek 1.14 *Maticová norma $\|\cdot\|_{a,b}$ je souhlasná s vektorovými normami $\|\cdot\|_a$ a $\|\cdot\|_b$, které ji indukují.*

Ze vztahu (1.5) rovněž plyne, že norma $\|\cdot\|_{a,b}$ je nejmenší maticová norma, která je souhlasná s vektorovými normami $\|\cdot\|_a$ a $\|\cdot\|_b$.

Nechť $A \in \mathbb{M}_{m,n}$ a $B \in \mathbb{M}_{n,p}$. Jestliže na \mathbb{R}^m , \mathbb{R}^n i \mathbb{R}^p zvolíme tutéž vektorovou normu $\|\cdot\|_a$, bude pro $\mathbf{x} \in \mathbb{R}^p$ platit, že $\|AB\mathbf{x}\|_a \leq \|A\|_{a,a} \cdot \|B\mathbf{x}\|_a \leq \|A\|_{a,a} \cdot \|B\|_{a,a} \cdot \|\mathbf{x}\|_a$. Protože $\|AB\|_a$ je nejmenší číslo k , které splňuje nerovnost $\|AB\mathbf{x}\|_a \leq k\|\mathbf{x}\|_a$, dostáváme, že $\|AB\|_a \leq \|A\|_{a,a} \cdot \|B\|_{a,a}$. Tudíž:

Důsledek 1.15 Maticová norma $\|\cdot\|_a$ indukovaná vektorovou normou $\|\cdot\|_a$ je submultiplikativní.

Nechť $\|\cdot\|_{a,b}$ je maticová norma indukovaná vektorovými normami $\|\cdot\|_a$ a $\|\cdot\|_b$. Pro $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_a \leq 1$, dostaneme, že $\|A\mathbf{x}\|_b \leq \|A\|_{a,b} \cdot \|\mathbf{x}\|_a \leq \|A\|_{a,b}$, odkud $\max_{\|\mathbf{x}\|_a \leq 1} \|A\mathbf{x}\|_b \leq \|A\|_{a,b}$. Na druhé straně, jelikož $S(\mathbf{0}, 1) \subseteq K(\mathbf{0}, 1)$, platí, že $\|A\|_{a,b} = \max_{\|\mathbf{x}\|_a=1} \|A\mathbf{x}\|_b \leq \max_{\|\mathbf{x}\|_a \leq 1} \|A\mathbf{x}\|_b$. Tedy $\|A\|_{a,b} = \max_{\|\mathbf{x}\|_a \leq 1} \|A\mathbf{x}\|_b$ a můžeme doplnit vztah (1.5) o třetí ekvivalentní definici. Maticovou normu indukovanou vektorovými normami můžeme tudíž definovat kterýmkoliv z následujících výrazů.

$$\|A\|_{a,b} = \max_{\|\mathbf{x}\|_a=1} \|A\mathbf{x}\|_b = \max_{\|\mathbf{x}\|_a \leq 1} \|A\mathbf{x}\|_b = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_b}{\|\mathbf{x}\|_a}. \quad (1.6)$$

Neužívanější operátorové maticové normy

Uvažujme matici $A \in \mathbb{M}_{m,n}$. Na množinách \mathbb{R}^m a \mathbb{R}^n budeme vždy uvažovat tutéž vektorovou normu $\|\cdot\|_p$, kde $p = 1, 2$ nebo ∞ . Půjde tedy o součtovou, eukleidovskou nebo maximální vektorovou normu. Lze ukázat, že maticové normy $\|\cdot\|_p$ indukované těmito vektorovými normami mají následující vzorce (viz např. [18, str. 167] nebo [52, str. 210]).

$$1) \|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^m |a_{ij}|.$$

Norma $\|A\|_1$ je tedy rovna maximu sloupcových součtů absolutních hodnot prvků matice A .

$$2) \|A\|_2 = \sqrt{\lambda_{\max}}, \text{ kde } \lambda_{\max} \text{ je největší vlastní číslo matice } A^T A.$$

$$3) \|A\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^n |a_{ij}|.$$

Norma $\|A\|_\infty$ je tedy rovna maximu řádkových součtů absolutních hodnot prvků matice A .

Vzorec pro maticovou normu $\|\cdot\|_2$ potřebuje jisté vysvětlení. Čtvercová matice B se nazývá *symetrická*, jestliže platí $B^T = B$. Je-li $B \in \mathbb{M}_n$ symetrická a pro libovolný nenulový vektor $\mathbf{x} \in \mathbb{R}^n$ platí $\mathbf{x}^T B \mathbf{x} > 0$ ($\mathbf{x}^T B \mathbf{x} \geq 0$), nazývá se matice B *pozitivně definitní (semidefinitní)*. Viz též str. 225. Uvědomte si, že $\mathbf{x}^T B \mathbf{x}$ je jednorozměrná matice, tj. číslo.

Lze dokázat, že všechna vlastní čísla, tj. kořeny charakteristického polynomu, pozitivně definitní (semidefinitní) matice jsou reálná a kladná (nezáporná). Matice $A^T A$ je symetrická, protože $(A^T A)^T = A^T (A^T)^T = A^T A$. Je rovněž pozitivně semidefinitní, protože $\mathbf{x}^T A^T A \mathbf{x} = (A \mathbf{x})^T A \mathbf{x} = \|A \mathbf{x}\|_2^2 \geq 0$. Tudíž $\lambda_{\max} \geq 0$ a zmíněný vzorec má smysl.

Podle důsledku 1.14 jsou všechny tři výše uvedené maticové normy souhlasné s normou, která je indukuje, tj. platí $\|A \mathbf{x}\|_p \leq \|A\|_p \cdot \|\mathbf{x}\|_p$, kde $p = 1, 2, \infty$. Podle důsledku 1.15 jsou tyto normy rovněž submultiplikativní, tj. platí $\|AB\|_p \leq \|A\|_p \cdot \|B\|_p$, kde $p = 1, 2, \infty$.

Příklad 1.16 Vypočítejte maticové normy $\|A\|_S$, $\|A\|_F$, $\|A\|_M$, $\|A\|_1$, $\|A\|_2$ a $\|A\|_\infty$ pro matici

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -3 & 1 & -2 \end{pmatrix}.$$

Řešení. Pomocí výše uvedených vzorců dostaneme:

$$\|A\|_S = |2| + |-1| + |0| + |-3| + |1| + |-2| = 9,$$

$$\|A\|_F = \sqrt{2^2 + (-1)^2 + 0^2 + (-3)^2 + 1^2 + (-2)^2} = \sqrt{19},$$

$$\|A\|_M = \max\{|2|, |-1|, |0|, |-3|, |1|, |-2|\} = 3,$$

$$\|A\|_1 = \max\{|2| + |-3|, |-1| + |1|, |0| + |-2|\} = 5,$$

$$\|A\|_\infty = \max\{|2| + |-1| + |0|, |-3| + |1| + |-2|\} = 6.$$

K určení normy $\|A\|_2$ musíme nejprve spočítat matici $A^T A$, najít její charakteristický

polynom a vypočítat jeho kořeny. Vyjde:

$$A^T A = \begin{pmatrix} 2 & -3 \\ -1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -3 & 1 & -2 \end{pmatrix} = \begin{pmatrix} 13 & -5 & 6 \\ -5 & 2 & -2 \\ 6 & -2 & 4 \end{pmatrix},$$

$$\det(A^T A - \lambda E) = \begin{vmatrix} 13 - \lambda & -5 & 6 \\ -5 & 2 - \lambda & -2 \\ 6 & -2 & 4 - \lambda \end{vmatrix} = -\lambda^3 + 19\lambda^2 - 21\lambda.$$

Jeden kořen je $\lambda_1 = 0$. Zbývající dva dostaneme řešením kvadratické rovnice $\lambda^2 - 19\lambda + 21 = 0$:

$$\lambda_{2,3} = \frac{19 \pm \sqrt{277}}{2}.$$

Tedy

$$\|A\|_2 = \sqrt{\lambda_{\max}} = \sqrt{\frac{19 + \sqrt{277}}{2}} \doteq 4,222.$$



Poznámka 1.17 Je-li $\|\cdot\|$ operátorová norma na \mathbb{M}_n indukovaná nějakou vektorovou normou na \mathbb{R}_n , je z její definice zřejmé, že $\|E\| = 1$, kde $E \in \mathbb{M}_n$ je jednotková matice. Protože $\|E\|_F = \sqrt{n}$, je jasné, že Frobeniova norma není pro $n > 1$ operátorová.

Poznámka 1.18 Výsledky tohoto oddílu platí i pro matice, jejichž prvky jsou komplexní čísla — srovnejte poznámku 1.10. Je pouze třeba provést několik drobných modifikací. Vektory budou prvky množiny \mathbb{C}^n .

V definici Frobeniovy normy je třeba změnit a_{ij}^2 (což nemusí být reálné číslo) na $|a_{ij}|^2$. Ve vzorci pro maticovou normu $\|\cdot\|_2$ je třeba matici $A^T A$ nahradit maticí $A^* A$. Zde A^* značí matici konjugovanou k A^T . Ta je rovna matici A^T , jejíž prvky se nahradí komplexně sdruženými čísly. Čtvercová matice $B = A^* A$ je pak tzv. *hermitovská*, což znamená, že pro ni platí $B^* = B$. Definice pozitivně definitní resp. semidefinitní matice se zavádí pro hermitovské matice. Pro nenulové vektory $\mathbf{x} \in \mathbb{C}^n$ musí platit $\mathbf{x}^* B \mathbf{x} > 0$ resp. $\mathbf{x}^* B \mathbf{x} \geq 0$, kde B je hermitovská matice. (Lze ověřit, že $\mathbf{x}^* B \mathbf{x}$ je pro hermitovskou matici reálné číslo.) Všechna vlastní čísla pozitivně definitních resp. semidefinitních matic jsou reálná a kladná resp. nezáporná.

1.7.3 Číslo podmíněnosti čtvercové matice

Předpokládejme, že $\|\cdot\|$ je maticová norma na množině čtvercových matic \mathbb{M}_n , která je submultiplikativní, je souhlasná s vektorovou normou $\|\cdot\|$ na \mathbb{R}^n a platí, že $\|E\| = 1$. Například to může být operátorová maticová norma indukovaná nějakou vektorovou normou — viz důsledky 1.14, 1.15 a poznámka 1.17.

Připomeňme, že čtvercová matice A se nazývá *regulární*, právě když $\det A \neq 0$. Tato vlastnost je současně ekvivalentní existenci inverzní matice A^{-1} .

Definice 1.19 Nechť A je regulární čtvercová matice. Číslo $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ se nazývá *číslo podmíněnosti matice A* .

Jak uvidíme v kapitole 3, číslo podmíněnosti matice má zásadní význam při vyšetřování citlivosti numerického řešení soustav lineárních algebraických rovnic na chybách ve vstupních datech a zaokrouhlovacích chybách.

Číslo podmíněnosti závisí pochopitelně na volbě konkrétní maticové normy. V aplikacích to však většinou nehraje podstatnou roli.

Vzhledem k předpokladům, které jsme udělali o použité maticové normě, platí

$$1 = \|E\| = \|AA^{-1}\| \leq \|A\| \cdot \|A^{-1}\| = \kappa(A).$$

Pro libovolnou čtvercovou matici je proto $\kappa(A) \geq 1$. Je-li $\kappa(A) \approx 1$, nazývá se matice *dobře podmíněná*. Je-li $\kappa(A)$ hodně větší než jedna, je matice *špatně podmíněná* — srovnejte definici podmíněnosti korektní úlohy na str. 30.

Na závěr uvedeme ještě jedno vyjádření čísla podmíněnosti matice. Je-li A regulární matice a vektor \mathbf{x} probíhá všechny nenulové vektory množiny \mathbb{R}^n , pak také vektory $A\mathbf{x}$ probíhají celou tuto množinu. Jak již bylo zmíněné před vztahem (1.4), funkce $\|A\mathbf{x}\|$ nabývá na $S(\mathbf{0}, 1)$ i minimum, které musí být pro regulární matici kladné, protože $A\mathbf{x} \neq \mathbf{0}$ pro $\mathbf{x} \neq \mathbf{0}$. Vypočítáme nyní normu inverzní matice. Vyjde:

$$\begin{aligned} \|A^{-1}\| &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A^{-1}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A^{-1}A\mathbf{x}\|}{\|A\mathbf{x}\|} = \\ &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{x}\|}{\|A\mathbf{x}\|} = \frac{1}{\min_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}} = \frac{1}{\min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|}. \end{aligned}$$

Dostáváme tak, že

$$\kappa(A) = \frac{\max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|}{\min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|} = \left(\max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \right) \cdot \left(\min_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \right)^{-1}.$$

Pro singulární matici je $\min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = 0$, proto je pro takovou matici rozumné položit $\kappa(A) = \infty$.

1.7.4 Spektrální poloměr

Uvažujme čtvercovou matici A s reálnými nebo komplexními prvky. Připomeňme, že vlastní čísla matice A jsou kořeny charakteristického polynomu $\det(A - \lambda E)$. Obecně jsou to komplexní čísla (i v případě, že matice A je reálná).

Definice 1.20 Necht' $\lambda_1, \lambda_2, \dots, \lambda_n$ jsou vlastní čísla matice A . Číslo

$$\rho(A) = \max_{i=1, \dots, n} |\lambda_i| \quad (1.7)$$

se nazývá *spektrální poloměr matice A* .

Spektrální poloměr hraje důležitou roli např. při vyšetřování konvergence iteračních metod řešení soustav lineárních rovnic. Určit jej je obecně poměrně obtížné, avšak jeho velikost lze odhadnout pomocí vhodné maticové normy.

V následující větě předpokládáme, že použitá maticová norma je souhlasná s nějakou vektorovou normou. Protože vlastní čísla mohou být komplexní, a tudíž složky příslušných vlastních vektorů mohou být rovněž komplexní, je třeba obecně pracovat s vektorovými

normami na \mathbb{C}^n a maticovými normami, které jsou určeny pro matice s komplexními prvky — viz poznámky 1.10 a 1.18.

Věta 1.21 *Nechť A je čtvercová matice. Pak pro její spektrální poloměr platí nerovnost $\rho(A) \leq \|A\|$, kde $\|\cdot\|$ je libovolná maticová norma, která je souhlasná s nějakou vektorovou normou.*

Důkaz. Nechť uvažovaná maticová norma je souhlasná s vektorovou normou $\|\cdot\|_a$. Předpokládejme, že λ je vlastní číslo matice A příslušné vlastnímu vektoru \mathbf{x} . Ze vztahu $A\mathbf{x} = \lambda\mathbf{x}$ dostaneme:

$$|\lambda| \cdot \|\mathbf{x}\|_a = \|\lambda\mathbf{x}\|_a = \|A\mathbf{x}\|_a \leq \|A\| \cdot \|\mathbf{x}\|_a.$$

Protože $\mathbf{x} \neq \mathbf{0}$, je $\|\mathbf{x}\|_a > 0$, takže z předchozí nerovnosti dostaneme, že $|\lambda| \leq \|A\|$. Jelikož vlastní číslo λ bylo libovolné, platí dokazovaná nerovnost. \square

1.7.5 Skalární součin

V oddílu 4.3 budou použity některé poznatky o ortogonálních vektorech v \mathbb{R}^n . Uvedeme nyní proto stručný přehled potřebných pojmů a výsledků.

Definice 1.22 Funkce, která každé dvojici vektorů $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ přiřazuje reálné číslo (\mathbf{x}, \mathbf{y}) , se nazývá *skalární součin na \mathbb{R}^n* , jestliže jsou splněny následující čtyři vlastnosti:

- 1) $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$ pro každé vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$;
- 2) $(c\mathbf{x}, \mathbf{y}) = c(\mathbf{x}, \mathbf{y})$ pro každé číslo $c \in \mathbb{R}$ a každé vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$;
- 3) $(\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z})$ pro každé vektory $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$;
- 4) $(\mathbf{x}, \mathbf{x}) > 0$ pro každý nenulový vektor $\mathbf{x} \in \mathbb{R}^n$.

První vlastnost se nazývá *komutativní zákon*, druhá *homogenita v první složce* a třetí *pravý distributivní zákon*. Poznamenejme, že v elementární analytické geometrii a ve fyzice se skalární součin obvykle značí tečkou, tj. $\mathbf{x} \cdot \mathbf{y}$. Naše označení má však své přednosti, zejména při zápisu složitějších výrazů.

Z definice skalárního součinu se snadno odvodí další vlastnosti, které pro něj platí.

a) $(\mathbf{x}, c\mathbf{y}) = (c\mathbf{y}, \mathbf{x}) = c(\mathbf{y}, \mathbf{x}) = c(\mathbf{x}, \mathbf{y})$ pro každé číslo $c \in \mathbb{R}$ a každé vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Platí tedy i homogenita ve druhé složce.

b) $(\mathbf{z}, \mathbf{x} + \mathbf{y}) = (\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z}) = (\mathbf{z}, \mathbf{x}) + (\mathbf{z}, \mathbf{y})$ pro každé vektory $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.

Platí tedy i levý distributivní zákon.

c) $(\mathbf{0}, \mathbf{x}) = (0 \cdot \mathbf{0}, \mathbf{x}) = 0(\mathbf{0}, \mathbf{x}) = 0$ pro každý vektor $\mathbf{x} \in \mathbb{R}^n$.

Tedy skalární součin nulového vektoru s libovolným dalším vektorem je číslo nula.

Čtvrtou vlastnost skalárního součinu lze proto zformulovat také tak, že $(\mathbf{x}, \mathbf{x}) \geq 0$ pro každý vektor $\mathbf{x} \in \mathbb{R}^n$, přičemž rovnost nastane pouze pro nulový vektor.

Skalární součin úzce souvisí s vektorovou normou, což je obsahem následující věty. Ta říká, že na prostoru se skalárním součinem je dána jistá přirozená norma.

Věta 1.23 *Nechť je na \mathbb{R}^n dán skalární součin. Pak funkce $\|\cdot\|$ daná vztahem*

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (1.8)$$

je vektorová norma na \mathbb{R}^n . Říkáme, že tato norma je indukovaná skalárním součinem.

Důkaz. První vlastnost normy je zřejmá. Ověření druhé vlastnosti normy je snadné. Pro $c \in \mathbb{R}$ a $\mathbf{x} \in \mathbb{R}^n$ s využitím homogenity v obou složkách vyjde

$$\|c\mathbf{x}\| = \sqrt{(c\mathbf{x}, c\mathbf{x})} = \sqrt{c^2(\mathbf{x}, \mathbf{x})} = \sqrt{c^2} \sqrt{(\mathbf{x}, \mathbf{x})} = |c| \cdot \|\mathbf{x}\|.$$

Třetí vlastnost, trojúhelníková nerovnost, se odvodí pomocí důležité nerovnosti, známé pod jmény *Cauchyova*¹-*Bunjakovského*²-*Schwarzova*³, která říká, že pro libovolné vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ platí

$$|(\mathbf{x}, \mathbf{y})| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|, \quad (1.9)$$

kde $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$ a $\|\mathbf{y}\| = \sqrt{(\mathbf{y}, \mathbf{y})}$. Důkaz viz např. [36, str. 66]. S využitím

¹**Augustin-Louis Cauchy** (1789–1857) (čti koši) — vynikající francouzský matematik, autor 789 prací. Položil základy současné matematiky, zejména analýzy.

²**Viktor Jakovlevič Bunjakovskij** (1804–1889) — ruský matematik. Zabýval se teorií čísel, geometrií a aplikovanou matematikou.

³**Hermann Amandus Schwarz** (1843–1921) (čti švarc) — německý matematik. Zabýval se analýzou a jejími aplikacemi v geometrii.

distributivního a komutativního zákona a této nerovnosti dostaneme:

$$\begin{aligned}\|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) = \\ &= \|\mathbf{x}\|^2 + 2(\mathbf{x}, \mathbf{y}) + \|\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.\end{aligned}$$

Po odmocnění obdržíme trojúhelníkovou nerovnost. □

Nejdůležitějším příkladem skalárního součinu na \mathbb{R}^n je tzv. *standardní skalární součin*, který je dán vzorcem

$$(\mathbf{x}, \mathbf{y}) = x_1y_1 + \cdots + x_ny_n.$$

Jím indukovaná vektorová norma je pak

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})} = \sqrt{x_1^2 + \cdots + x_n^2},$$

což je nám již známá eukleidovská norma, kterou značíme $\|\mathbf{x}\|_2$.

Není těžké popsat všechny skalární součiny na \mathbb{R}^n . Protože však v dalším textu budeme potřebovat právě standardní skalární součin, další příklady nebudeme uvádět. Nicméně následující výsledky platí pro libovolné skalární součiny.

Poznámka 1.24 Viděli jsme, že eukleidovskou vektorovou normu lze indukovat skalárním součinem. Je přirozené položit si otázku, zda každá vektorová norma může být indukovaná nějakým skalárním součinem. Odpověď je negativní.

Pomocí vlastností skalárního součinu se snadno ověří, že pro normu indukovanou skalárním součinem, tj. danou vztahem (1.8), platí identita

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2.$$

Lze ukázat, že tato podmínka je i postačující. Důkaz není triviální, viz např. [32, str. 68].

Pravoúhlý průmět vektoru na podprostor

V oddílu 4.3 budeme potřebovat umět rozložit daný vektor na součet dvou kolmých vektorů. Zavést pojem kolmosti nám umožní právě skalární součin.

Definice 1.25 Předpokládejme, že na \mathbb{R}^n je dán skalární součin (\cdot, \cdot) .

Řekneme, že dva vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ jsou *ortogonální* neboli *kolmé* (vzhledem k tomuto skalárnímu součinu) a píšeme $\mathbf{x} \perp \mathbf{y}$, jestliže platí

$$(\mathbf{x}, \mathbf{y}) = 0. \quad (1.10)$$

Dále řekneme, že vektor \mathbf{x} je *ortogonální* nebo *kolmý* k množině vektorů $A \subset \mathbb{R}^n$ a píšeme $\mathbf{x} \perp A$, jestliže je vektor \mathbf{x} ortogonální ke každému vektoru $\mathbf{y} \in A$.

V případě množin \mathbb{R}^2 a \mathbb{R}^3 se standardním skalárním součinem, jejichž geometrickými modely jsou rovina a (trojrozměrný) prostor s eukleidovskou vzdáleností, jde o obvyklý pojem kolmosti, jak ho známe z analytické geometrie.

Nechť $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)} \in \mathbb{R}^n$ je daná množina m vektorů. Uvažujme nyní mno-

žinu $U \subset \mathbb{R}^n$ všech vektorů $\mathbf{x} \in \mathbb{R}^n$ tvaru

$$\mathbf{x} = c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)} + \dots + c_m \mathbf{x}^{(m)}, \quad (1.11)$$

kde c_1, c_2, \dots, c_m jsou libovolná reálná čísla. Říkáme, že vektor \mathbf{x} je *lineární kombinací vektorů* $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$ s *koefficienty* c_1, \dots, c_m . Množina U se nazývá *podprostor* \mathbb{R}^n *generovaný vektory* $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$ a použijeme pro ni označení

$$U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}].$$

Snadno je vidět, že pokud $\mathbf{x}, \mathbf{y} \in U$ a c je libovolné číslo, pak také $\mathbf{x} + \mathbf{y} \in U$ a $c\mathbf{x} \in U$. Vektory \mathbf{x} a \mathbf{y} mají totiž vyjádření $\mathbf{x} = c_1 \mathbf{x}^{(1)} + \dots + c_m \mathbf{x}^{(m)}$ a $\mathbf{y} = d_1 \mathbf{x}^{(1)} + \dots + d_m \mathbf{x}^{(m)}$, takže

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (c_1 + d_1) \mathbf{x}^{(1)} + \dots + (c_m + d_m) \mathbf{x}^{(m)}, \\ c\mathbf{x} &= cc_1 \mathbf{x}^{(1)} + \dots + cc_m \mathbf{x}^{(m)}, \end{aligned}$$

což jsou opět vektory mající vyjádření ve tvaru (1.11). Tedy množina U s každými dvěma vektory obsahuje jejich součet a s každým vektorem obsahuje jeho libovolný násobek.

Právě podmnožiny \mathbb{R}^n mající tyto dvě vlastnosti se nazývají podprostory \mathbb{R}^n . Všimněte si, že zejména $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)} \in U$. Pro libovolné $i = 1, \dots, m$ stačí položit $c_i = 1$ a $c_j = 0$ pro $i \neq j$. Pak $c_1 \mathbf{x}^{(1)} + \dots + c_m \mathbf{x}^{(m)} = \mathbf{x}^{(i)}$.

Dále uvažujme vektory $\mathbf{x} \in \mathbb{R}^n$, které jsou ortogonální k množině U , tj. $\mathbf{x} \perp U$. Množinu všech takových vektorů nazýváme *ortogonální doplněk* U a značíme ji U^\perp . Tedy

$$\mathbf{x} \in U^\perp, \text{ právě když } (\mathbf{x}, \mathbf{y}) = 0 \text{ pro každý vektor } \mathbf{y} \in U.$$

Pomocí vlastností skalárního součinu (distributivního zákona a homogenity) se snadno ověří, že pokud $\mathbf{x}, \mathbf{y} \in U^\perp$ a c je libovolné číslo, pak také $\mathbf{x} + \mathbf{y} \in U^\perp$ a $c\mathbf{x} \in U^\perp$. Pro libovolný vektor $\mathbf{z} \in U$ totiž platí:

$$\begin{aligned}(\mathbf{x} + \mathbf{y}, \mathbf{z}) &= (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z}) = 0 + 0 = 0, \\(c\mathbf{x}, \mathbf{z}) &= c(\mathbf{x}, \mathbf{z}) = c \cdot 0 = 0.\end{aligned}$$

Tudíž ortogonální doplněk U^\perp je také podprostor \mathbb{R}^n .

Protože všechny vektory $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$ leží v U , musí k nim být libovolný vektor $\mathbf{y} \in U^\perp$ ortogonální, tj. musí platit, že $\mathbf{y} \perp \mathbf{x}^{(i)}$, $i = 1, \dots, m$. Platí však i opačné tvrzení, tzv. *kritérium kolmosti*: Je-li některý vektor $\mathbf{y} \in \mathbb{R}^n$ ortogonální ke všem vektorům

$\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$, pak $\mathbf{y} \in U^\perp$. Skutečně. Libovolný vektor $\mathbf{x} \in U$ má tvar (1.11). Tedy

$$\begin{aligned}(\mathbf{y}, \mathbf{x}) &= (\mathbf{y}, c_1 \mathbf{x}^{(1)} + \dots + c_m \mathbf{x}^{(m)}) = c_1 (\mathbf{y}, \mathbf{x}^{(1)}) + \dots + c_m (\mathbf{y}, \mathbf{x}^{(m)}) = \\ &= c_1 \cdot 0 + \dots + c_m \cdot 0 = 0,\end{aligned}$$

takže vektor \mathbf{y} je kolmý ke každému vektoru z podprostoru U .

Nyní již můžeme zformulovat hlavní výsledek.

Věta 1.26 *Nechť $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}]$ je podprostor \mathbb{R}^n a U^\perp je jeho ortogonální doplněk.*

K libovolnému vektoru $\mathbf{y} \in \mathbb{R}^n$ existuje právě jedna dvojice vektorů $\mathbf{x} \in U$ a $\mathbf{z} \in U^\perp$ taková, že $\mathbf{y} = \mathbf{x} + \mathbf{z}$.

Důkaz viz např. [36, str. 96]. Vektor \mathbf{x} se nazývá *pravoúhlý průmět vektoru \mathbf{y} na podprostor U* . Věta říká, že libovolný vektor lze jednoznačně rozložit na součet ortogonálních složek, přičemž jedna z nich leží v daném podprostoru.

Pravoúhlý průmět \mathbf{x} má důležitou vlastnost. Je nejlepší aproximací vektoru \mathbf{y} ze všech vektorů $\hat{\mathbf{x}} \in U$, tj. minimalizuje výraz $\|\mathbf{y} - \hat{\mathbf{x}}\|$. Skutečně. Platí totiž, že $\mathbf{y} - \hat{\mathbf{x}} = \mathbf{x} + \mathbf{z} - \hat{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}} + \mathbf{z}$, kde $\mathbf{x} - \hat{\mathbf{x}} \in U$ a $\mathbf{z} \in U^\perp$, takže $(\mathbf{x} - \hat{\mathbf{x}}, \mathbf{z}) = 0$.

Odtud

$$\begin{aligned}\|y - \hat{x}\|^2 &= \|x + z - \hat{x}\|^2 = (x - \hat{x} + z, x - \hat{x} + z) = (x - \hat{x}, x - \hat{x}) + \\ &+ (x - \hat{x}, z) + (z, x - \hat{x}) + (z, z) = \|x - \hat{x}\|^2 + \|z\|^2.\end{aligned}$$

Pro $x \neq \hat{x}$ je proto

$$\|y - \hat{x}\|^2 = \|x - \hat{x}\|^2 + \|z\|^2 > \|z\|^2 = \|y - x\|^2.$$

Tudíž

$$\|y - x\| = \min_{\hat{x} \in U} \|y - \hat{x}\|, \quad (1.12)$$

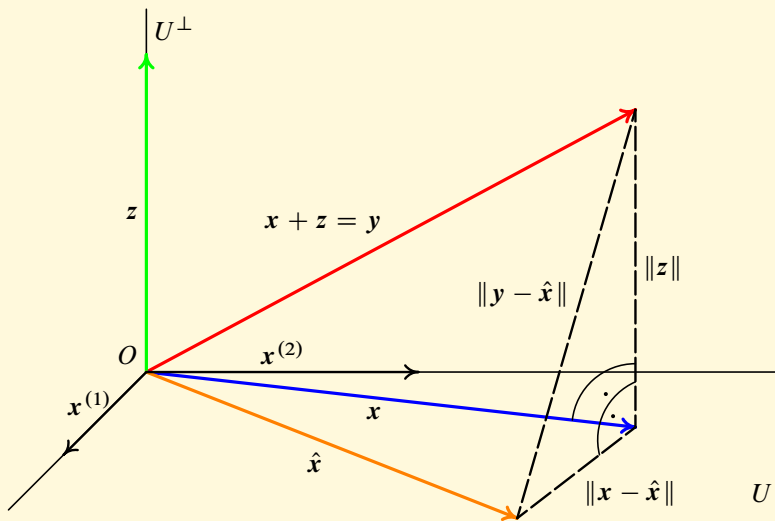
přičemž pravouhlý průmět x je jediný vektor, v něž se toto minimum nabývá. Dále je vidět, že číslo $\|z\|$ udává, jaké chyby se při náhradě vektoru y vektorem x dopustíme.

Pro lepší představu znázorníme pravouhlý průmět vektoru y v množině \mathbb{R}^3 se standardním skalárním součinem, která je modelem trojrozměrného prostoru s běžnou eukleidovskou metrikou. Kolmost má tedy běžný význam a vše si dokážeme představit. Vektory umístíme do počátku O . Trojici čísel chápeme jako souřadnice koncového bodu vázaného vektoru, tj. orientované úsečky, jejímž počátečním bodem je počátek.

Nechť $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$ a předpokládejme, že vektory $\mathbf{x}^{(1)}$ a $\mathbf{x}^{(2)}$ jsou nekolineární (jeden není násobkem druhého). Pak trojice tvaru $c_1\mathbf{x}^{(1)} + c_2\mathbf{x}^{(2)}$, kde c_1 a c_2 jsou libovolná čísla, odpovídají bodům, které leží v rovině procházející počátkem a jejímiž směrovými vektory jsou právě $\mathbf{x}^{(1)}$ a $\mathbf{x}^{(2)}$. Vlastně jde o parametrické rovnice této roviny.

Zvolme pro jednoduchost $\mathbf{x}^{(1)} = [1, 0, 0]^T$ a $\mathbf{x}^{(2)} = [0, 1, 0]^T$ (v analytické geometrii tyto vektory obvykle značíme \mathbf{i} a \mathbf{j}). Potom podprostor U splývá s půdorysnou a ortogonální doplněk U^\perp splývá s přímkou, která je k půdorysně kolmá a prochází počátkem — viz obr. 1.3.

Nechť vektor \mathbf{y} neleží v podprostoru U . Jeho pravoúhlý průmět \mathbf{x} dostaneme tak, že z koncového bodu spustíme kolmici na půdorysnu. Dále $\mathbf{z} = \mathbf{y} - \mathbf{x}$ a tento vektor leží v U^\perp . Zvolme vektor $\hat{\mathbf{x}} \in U$, $\hat{\mathbf{x}} \neq \mathbf{x}$. Pak koncové body vektorů \mathbf{y} , \mathbf{x} a $\hat{\mathbf{x}}$ určují pravoúhlý trojúhelník. Délky jeho stran jsou $\|\mathbf{y} - \hat{\mathbf{x}}\|$ (přepona), $\|\mathbf{x} - \hat{\mathbf{x}}\|$ a $\|\mathbf{z}\|$ (odvěsny). Vztah $\|\mathbf{y} - \hat{\mathbf{x}}\|^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \|\mathbf{z}\|^2$ tedy představuje Pythagorovu větu. Z obrázku je zřejmé, že délka $\|\mathbf{y} - \hat{\mathbf{x}}\|$ je minimální, právě když je $\hat{\mathbf{x}} = \mathbf{x}$.

Obr. 1.3: Pravoúhlý průmět vektoru v \mathbb{R}^3

Nalezení pravouhlého průmětu

Na závěr popíšeme, jak lze složky jednoznačného rozkladu $\mathbf{y} = \mathbf{x} + \mathbf{z}$ určit. Předpokládejme, že $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}]$. Protože $\mathbf{x} \in U$, lze tento vektor vyjádřit ve tvaru $\mathbf{x} = c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)} + \dots + c_m \mathbf{x}^{(m)}$. Po dosazení dostaneme

$$\mathbf{y} = c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)} + \dots + c_m \mathbf{x}^{(m)} + \mathbf{z}.$$

V předchozí rovnici neznáme konstanty c_1, c_2, \dots, c_m a vektor \mathbf{z} . Toho se však elegantně zbavíme. Víme totiž, že leží v U^\perp , takže je ortogonální ke všem vektorům $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ až $\mathbf{x}^{(m)}$, tudíž $(\mathbf{x}^{(i)}, \mathbf{z}) = 0, i = 1, \dots, m$. Předchozí rovnost proto postupně těmito vektory skalárně vynásobíme. Dostaneme:

$$\begin{aligned}
(\mathbf{x}^{(1)}, \mathbf{y}) &= c_1(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) + c_2(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) + \dots + c_m(\mathbf{x}^{(1)}, \mathbf{x}^{(m)}), \\
(\mathbf{x}^{(2)}, \mathbf{y}) &= c_1(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) + c_2(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) + \dots + c_m(\mathbf{x}^{(2)}, \mathbf{x}^{(m)}), \\
&\vdots \\
(\mathbf{x}^{(m-1)}, \mathbf{y}) &= c_1(\mathbf{x}^{(m-1)}, \mathbf{x}^{(1)}) + c_2(\mathbf{x}^{(m-1)}, \mathbf{x}^{(2)}) + \dots + c_m(\mathbf{x}^{(m-1)}, \mathbf{x}^{(m)}), \\
(\mathbf{x}^{(m)}, \mathbf{y}) &= c_1(\mathbf{x}^{(m)}, \mathbf{x}^{(1)}) + c_2(\mathbf{x}^{(m)}, \mathbf{x}^{(2)}) + \dots + c_m(\mathbf{x}^{(m)}, \mathbf{x}^{(m)}).
\end{aligned} \tag{1.13}$$

To je soustava m lineárních rovnic pro m neznámých. Nazývá se *normální soustava rovnic*. Věta 1.26 zaručuje, že tato soustava má řešení. Jakmile známe konstanty c_1, c_2, \dots, c_m , můžeme určit vektor $\mathbf{x} = c_1\mathbf{x}^{(1)} + c_2\mathbf{x}^{(2)} + \dots + c_m\mathbf{x}^{(m)}$ a následně i vektor $\mathbf{z} = \mathbf{y} - \mathbf{x}$.

Na závěr si všimneme, kolik má soustava (1.13) řešení. Matice soustavy má tvar

$$G(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}) = \begin{pmatrix} (\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & (\mathbf{x}^{(1)}, \mathbf{x}^{(m)}) \\ (\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & (\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & (\mathbf{x}^{(2)}, \mathbf{x}^{(m)}) \\ \dots & \dots & \dots & \dots \\ (\mathbf{x}^{(m-1)}, \mathbf{x}^{(1)}) & (\mathbf{x}^{(m-1)}, \mathbf{x}^{(2)}) & \dots & (\mathbf{x}^{(m-1)}, \mathbf{x}^{(m)}) \\ (\mathbf{x}^{(m)}, \mathbf{x}^{(1)}) & (\mathbf{x}^{(m)}, \mathbf{x}^{(2)}) & \dots & (\mathbf{x}^{(m)}, \mathbf{x}^{(m)}) \end{pmatrix}.$$

Nazývá se Gramova¹ matice systému vektorů $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$. Je symetrická a lze o ní dokázat, že $\det G(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}) \geq 0$, přičemž rovnost nastane právě tehdy, když jsou vektory $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$ lineárně závislé, tj. když je hodnota matice, která má za sloupce právě tyto vektory, menší než m .

Pokud jsou tedy vektory $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}$, které generují podprostor U , lineárně nezávislé, je matice soustavy regulární a soustava (1.13) má jediné řešení. Jsou-li závislé, má soustava nekonečně mnoho řešení závislých na jednom nebo více parametrech, ale díky jednoznačnosti pravoúhlého průmětu vyjde tento vektor vždy stejný (parametry se po dosazení za c_1, \dots, c_m do vzorce pro \mathbf{x} vyruší).

Příklad 1.27 V prostoru \mathbb{R}^3 se standardním skalárním součinem jsou dány vektory $\mathbf{x}^{(1)} = [1, -1, 2]^T$, $\mathbf{x}^{(2)} = [-3, 1, -2]^T$ a $\mathbf{y} = [3, -1, 1]^T$. Najděte pravoúhlý průmět vektoru \mathbf{y} na podprostor $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$.

Řešení. Podle věty 1.26 lze vektor \mathbf{y} vyjádřit ve tvaru $\mathbf{y} = \mathbf{x} + \mathbf{z}$, kde $\mathbf{x} \in U$ a $\mathbf{z} \in U^\perp$. Tedy $\mathbf{x} = c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)}$ a

$$\mathbf{y} = c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)} + \mathbf{z}.$$

¹Jørgen Pedersen Gram (1850–1916) — dánský matematik.

Předchozí rovnost skalárně vynásobíme postupně vektory $\mathbf{x}^{(1)}$ a $\mathbf{x}^{(2)}$. Dostaneme:

$$(\mathbf{x}^{(1)}, \mathbf{y}) = c_1(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) + c_2(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) + (\mathbf{x}^{(1)}, \mathbf{z}),$$

$$(\mathbf{x}^{(2)}, \mathbf{y}) = c_1(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) + c_2(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) + (\mathbf{x}^{(2)}, \mathbf{z}).$$

Víme, že $(\mathbf{x}^{(1)}, \mathbf{z}) = (\mathbf{x}^{(2)}, \mathbf{z}) = 0$. Vypočítáme zbývající skalární součiny.

$$(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) = 6, \quad (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = -8, \quad (\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) = 14,$$

$$(\mathbf{x}^{(1)}, \mathbf{y}) = 6, \quad (\mathbf{x}^{(2)}, \mathbf{y}) = -12.$$

Soustava lineárních rovnic má tvar

$$\begin{aligned} 6c_1 - 8c_2 &= 6, \\ -8c_1 + 14c_2 &= -12. \end{aligned}$$

Jejím jediným řešením je

$$c_1 = -\frac{3}{5}, \quad c_2 = -\frac{6}{5}.$$

Po dosazení máme

$$\mathbf{x} = -\frac{3}{5} [1, -1, 2]^T - \frac{6}{5} [-3, 1, -2]^T = \left[3, -\frac{3}{5}, \frac{6}{5} \right]^T$$

a

$$\mathbf{z} = \mathbf{y} - \mathbf{x} = [3, -1, 1]^T - \left[3, -\frac{3}{5}, \frac{6}{5} \right]^T = \left[0, -\frac{2}{5}, -\frac{1}{5} \right]^T.$$

Podle (1.12) pak

$$\|\mathbf{y} - \mathbf{x}\| = \min_{\hat{\mathbf{x}} \in U} \|\mathbf{y} - \hat{\mathbf{x}}\| = \|\mathbf{z}\| = \frac{1}{\sqrt{5}}$$

je nejmenší chyba, které lze dosáhnout při náhradě vektoru \mathbf{y} vektorem $\mathbf{z} \in U$. ▲

Poznámka 1.28 Skalární součin lze zavést i pro vektory z \mathbb{C}^n , jejichž složky jsou komplexní čísla. Skalární součin (\mathbf{x}, \mathbf{y}) , kde $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, bude komplexní číslo.

V definici 1.22 je třeba udělat jedinou změnu a to ve vlastnosti 1). Místo komutativního zákona musí platit, že $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$, kde pruh značí komplexně sdružené číslo. Vše ostatní je stejné. Vlastnost 4) navíc říká, že (\mathbf{x}, \mathbf{x}) je vždy reálné číslo. Místo homogenity ve druhé složce dostaneme, že platí $(\mathbf{x}, c\mathbf{y}) = \bar{c}(\mathbf{x}, \mathbf{y})$.

Standardní skalární součin se definuje vztahem $(\mathbf{x}, \mathbf{y}) = x_1 \bar{y}_1 + \dots + x_n \bar{y}_n$.

Pojmy k zapamatování

- matematický model
- numerická úloha
- numerický algoritmus
- chyba numerické úlohy
- zaokrouhlovací chyby
- absolutní a relativní chyba aproximace
- platná cifra a platné desetinné místo
- strojová čísla
- korektní úloha
- číslo podmíněnosti úlohy
- dobře podmíněný a numericky stabilní algoritmus
- symbol O
- vektorová norma

- maticová norma
- operátorová norma matice
- číslo podmíněnosti matice
- spektrální poloměr matice
- skalární součin
- pravoúhlý průmět vektoru na podprostor

Kontrolní otázky

1. Čím se zabývá numerická matematika?
2. Popište jednotlivé kroky postupu od reálného problému k numerickému algoritmu.
3. S jakými typy chyb se můžeme setkat při numerickém řešení úloh?
4. Vysvětlete pojmy aproximace čísla, absolutní a relativní chyby aproximace.
5. Co je to platná cifra a platné desetinné místo?
6. Jak jsou reprezentována čísla v počítači a co jsou to strojová čísla?
7. Co je příčinou zaokrouhlovacích chyb, když provádíme výpočty na počítači?

8. Co znamená, že numerická úloha je korektní?
9. Co znamená, že numerická úloha je dobře podmíněna?
10. Jak je definováno číslo podmíněnosti korektní úlohy?
11. Co znamená, že numerický algoritmus je stabilní?
12. Co znamená zápis $f(x) = O(g(x))$? Vysvětlete např. pro funkci $g(x) = x^4$ (načrtněte obrázek).
13. Jak je definovaná norma vektoru? Co je výsledkem normy vektoru?
14. Uveďte příklady vektorových norem.
15. Co jsou to ekvivalentní vektorové normy?
16. Jak je definovaná konvergence posloupnosti vektorů?
17. Jak je definovaná norma matice? Co je výsledkem normy matice?
18. Uveďte příklady maticových norem.
19. Co znamená, že maticová norma je souhlasná s vektorovými normami?
20. Co je to submultiplikativní maticová norma?
21. Uveďte vztah, kterým je definovaná maticová norma indukovaná vektorovými normami.
22. Uveďte příklady operátorových maticových norem.

23. Jak je definované číslo podmíněnosti matice?
24. Co je to spektrální poloměr matice?
25. Jak je definovaný skalární součin vektorů? Co je výsledkem skalárního součinu?
26. Jak se zavádí norma indukovaná skalárním součinem?
27. Jak je definován standardní skalární součin vektorů v \mathbb{R}^n ?
28. Kdy říkáme, že jsou dva vektory ortogonální?
29. Vysvětlete, co je to pravoúhlý průmět vektoru na podprostor. Vysvětlete s pomocí obrázku v \mathbb{R}^3 .
30. Popište postup nalezení pravoúhlého průmětu.

Cvičení

1. Vypočtete součtovou, eukleidovskou a maximální vektorovou normu vektorů:

$$\mathbf{x} = [-3, 2, -7]^T \in \mathbb{R}^3, \mathbf{y} = [-5, 2, 1, 3]^T \in \mathbb{R}^4, \mathbf{z} = [-5, 13, -2, 7, 3]^T \in \mathbb{R}^5, \\ \mathbf{u} = [-1, 3, -4, 8, -3, -7]^T \in \mathbb{R}^6.$$

2. Pro dané vektory vypočtěte normy $\|\cdot\|_1$, $\|\cdot\|_2$ a $\|\cdot\|_\infty$:

$$\mathbf{x} = [4, 2, -1, 8]^T \in \mathbb{R}^4, \mathbf{y} = [-5, 3, 0, 1, 4, -2, 9]^T \in \mathbb{R}^7.$$

3. Pro dané vektory vypočtěte normy $\|\cdot\|_1$, $\|\cdot\|_2$ a $\|\cdot\|_\infty$:

$$\mathbf{x} = [2, 3, -4, -7, 1]^T \in \mathbb{R}^5, \mathbf{y} = [-1, -2, -4, -3, 1, 2, 1]^T \in \mathbb{R}^7.$$

4. Pro dané vektory vypočtěte normy $\|\cdot\|_1$, $\|\cdot\|_2$ a $\|\cdot\|_\infty$:

$$\mathbf{x} = [-1, -9, -14, 24, -31, -7]^T \in \mathbb{R}^6, \mathbf{y} = [-11, 2, 44, -83]^T \in \mathbb{R}^4.$$

5. Vypočtěte součtovou, Frobeniovu a maximální normu následujících matic.

$$A = \begin{pmatrix} 2 & 2 & 3 \\ -1 & 4 & 5 \\ -3 & 2 & 7 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 4 & 5 & -3 \\ -2 & 0 & -3 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & -2 \\ -3 & 2 \\ -8 & 7 \\ -11 & 2 \end{pmatrix}.$$

6. Vypočtěte maticové normy $\|\cdot\|_S$, $\|\cdot\|_F$, $\|\cdot\|_M$, $\|\cdot\|_1$, $\|\cdot\|_2$ a $\|\cdot\|_\infty$ pro matice

$$A = \begin{pmatrix} -1 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 4 \\ 5 & -3 \end{pmatrix}, \quad C = \begin{pmatrix} -2 & -5 \\ 5 & 2 \end{pmatrix}.$$

7. Vypočtete maticové normy $\|\cdot\|_S$, $\|\cdot\|_F$, $\|\cdot\|_M$, $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ pro matice

$$A = \begin{pmatrix} -1 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix}, \quad B = A^T, \quad C = \begin{pmatrix} 3 & 1 & -5 \\ 1 & 3 & 1 \end{pmatrix}, \quad D = C^T.$$

8. Vypočtete maticové normy $\|\cdot\|_S$, $\|\cdot\|_F$, $\|\cdot\|_M$, $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ pro matice

$$A = \begin{pmatrix} -1 & 1 \\ 2 & -1 \\ 4 & -3 \end{pmatrix}, \quad B = A^T, \quad C = \begin{pmatrix} 1 & -2 \\ 5 & -1 \\ 2 & -1 \end{pmatrix}, \quad D = C^T.$$

9. Vypočtete maticové normy $\|\cdot\|_S$, $\|\cdot\|_F$, $\|\cdot\|_M$, $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ následujících matic. Řešení pro $\|\cdot\|_2$ je třeba získat numericky.

$$A = \begin{pmatrix} 3 & -5 & -4 \\ 2 & -2 & 1 \\ 0 & -3 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & -4 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} -3 & 5 & 1 \\ 2 & 0 & 1 \\ -2 & 0 & 2 \end{pmatrix}.$$

10. Určete spektrální poloměr čtvercových matic.

$$A = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & -4 & 2 \\ -4 & 8 & -4 \\ 2 & -4 & 11 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 2 \\ 5 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix},$$

$$F = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}, \quad G = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix},$$

$$J = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}, \quad K = \begin{pmatrix} 1 & -2 \\ 5 & -1 \end{pmatrix}, \quad L = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}, \quad M = \begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix}.$$

11. V prostoru \mathbb{R}^3 se standardním skalárním součinem jsou dány vektory $\mathbf{x}^{(1)} = [-1, 1, 1]^T$, $\mathbf{x}^{(2)} = [1, -1, 1]^T$ a $\mathbf{y} = [3, -1, 7]^T$. Najděte pravoúhlý průmět \mathbf{z} vektoru \mathbf{y} na podprostor $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$.

12. V prostoru \mathbb{R}^3 se standardním skalárním součinem jsou dány vektory $\mathbf{x}^{(1)} = [1, 1, 1]^T$, $\mathbf{x}^{(2)} = [-3, 3, 6]^T$ a $\mathbf{y} = [2, 1, 4]^T$. Najděte pravoúhlý průmět \mathbf{z} vektoru \mathbf{y} na podprostor $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$.
13. V prostoru \mathbb{R}^4 se standardním skalárním součinem jsou dány vektory $\mathbf{x}^{(1)} = [1, -1, 0, 2]^T$, $\mathbf{x}^{(2)} = [1, 2, 1, 1]^T$ a $\mathbf{y} = [1, -2, 3, 4]^T$. Najděte pravoúhlý průmět \mathbf{z} vektoru \mathbf{y} na podprostor $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$.
14. V prostoru \mathbb{R}^4 se standardním skalárním součinem jsou dány vektory $\mathbf{x}^{(1)} = [1, 1, 1, 1]^T$, $\mathbf{x}^{(2)} = [-2, -1, 3, 2]^T$ a $\mathbf{y} = [2, -2, -1, 1]^T$. Najděte pravoúhlý průmět \mathbf{z} vektoru \mathbf{y} na podprostor $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$.
15. V prostoru \mathbb{R}^4 se standardním skalárním součinem jsou dány vektory $\mathbf{x}^{(1)} = [1, 1, 1, 1]^T$, $\mathbf{x}^{(2)} = [-2, -1, 0, 1]^T$, $\mathbf{x}^{(3)} = [4, 1, 0, 1]^T$ a $\mathbf{y} = [-1, 2, 0, 1]^T$. Najděte pravoúhlý průmět \mathbf{z} vektoru \mathbf{y} na podprostor $U = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}]$.

Řešení

1. $\|\mathbf{x}\|_1 = 12$, $\|\mathbf{x}\|_2 = \sqrt{62}$, $\|\mathbf{x}\|_\infty = 7$, $\|\mathbf{y}\|_1 = 11$, $\|\mathbf{y}\|_2 = \sqrt{39}$, $\|\mathbf{y}\|_\infty = 5$,
 $\|\mathbf{z}\|_1 = 30$, $\|\mathbf{z}\|_2 = 16$, $\|\mathbf{z}\|_\infty = 39$, $\|\mathbf{u}\|_1 = 26$, $\|\mathbf{u}\|_2 = 2\sqrt{37}$, $\|\mathbf{u}\|_\infty = 8$.
2. $\|\mathbf{x}\|_1 = 15$, $\|\mathbf{x}\|_2 = \sqrt{85}$, $\|\mathbf{x}\|_\infty = 8$, $\|\mathbf{y}\|_1 = 24$, $\|\mathbf{y}\|_2 = 2\sqrt{34}$, $\|\mathbf{y}\|_\infty = 9$.
3. $\|\mathbf{x}\|_1 = 17$, $\|\mathbf{x}\|_2 = \sqrt{79}$, $\|\mathbf{x}\|_\infty = 7$, $\|\mathbf{y}\|_1 = 14$, $\|\mathbf{y}\|_2 = 6$, $\|\mathbf{y}\|_\infty = 4$.
4. $\|\mathbf{x}\|_1 = 86$, $\|\mathbf{x}\|_2 = 2\sqrt{466}$, $\|\mathbf{x}\|_\infty = 31$, $\|\mathbf{y}\|_1 = 140$, $\|\mathbf{y}\|_2 = 5\sqrt{358}$, $\|\mathbf{y}\|_\infty = 83$.
5. $\|A\|_S = 29$, $\|A\|_F = 11$, $\|A\|_M = 7$, $\|B\|_S = 29$, $\|B\|_F = s\sqrt{17}$, $\|B\|_M = 5$,
 $\|C\|_S = 36$, $\|C\|_F = 16$, $\|C\|_M = 11$.
6. $\|A\|_S = 10$, $\|A\|_F = \sqrt{14}$, $\|A\|_M = 2$, $\|A\|_1 = 4$, $\|A\|_2 = \sqrt{4 + \sqrt{10}}$, $\|A\|_\infty = 4$,
 $\|B\|_S = 14$, $\|B\|_F = 3\sqrt{6}$, $\|B\|_M = 5$, $\|B\|_1 = 7$, $\|B\|_2 = \sqrt{27 + \sqrt{53}}$, $\|B\|_\infty = 8$,
 $\|C\|_S = 14$, $\|C\|_F = \sqrt{58}$, $\|C\|_M = 5$, $\|C\|_1 = 7$, $\|C\|_2 = 7$, $\|C\|_\infty = 7$.
7. $\|A\|_S = 7$, $\|A\|_F = \sqrt{11}$, $\|A\|_M = 2$, $\|A\|_1 = 3$, $\|A\|_2 = \sqrt{6}$, $\|A\|_\infty = 4$, $\|B\|_S = 7$,
 $\|B\|_F = \sqrt{11}$, $\|B\|_M = 2$, $\|B\|_1 = 4$, $\|B\|_2 = \sqrt{6}$, $\|B\|_\infty = 3$, $\|C\|_S = 14$, $\|C\|_F =$

$$= \sqrt{46}, \|C\|_M = 5, \|C\|_1 = 6, \|C\|_2 = \sqrt{23 + \sqrt{145}}, \|C\|_\infty = 9, \|D\|_S = 14, \\ \|D\|_F = \sqrt{46}, \|D\|_M = 5, \|D\|_1 = 9, \|D\|_2 = \sqrt{23 + \sqrt{145}}, \|D\|_\infty = 6.$$

$$8. \|A\|_S = 12, \|A\|_F = 4\sqrt{2}, \|A\|_M = 4, \|A\|_1 = 7, \|A\|_2 = \sqrt{16 + 5\sqrt{10}}, \|A\|_\infty = 7, \\ \|B\|_S = 12, \|B\|_F = 4\sqrt{2}, \|B\|_M = 4, \|B\|_1 = 7, \|B\|_2 = \sqrt{16 + 5\sqrt{10}}, \|B\|_\infty = \\ = 7, \|C\|_S = 12, \|C\|_F = 6, \|C\|_M = 5, \|C\|_1 = 8, \|C\|_2 = \sqrt{33}, \|C\|_\infty = 6, \\ \|D\|_S = 12, \|D\|_F = 6, \|D\|_M = 5, \|D\|_1 = 6, \|D\|_2 = \sqrt{33}, \|D\|_\infty = 8.$$

$$9. \|A\|_S = 26, \|A\|_F = 2\sqrt{26}, \|A\|_M = 6, \|A\|_1 = 11, \|A\|_2 = \sqrt{57}, \|A\|_\infty = 12, \\ \|B\|_S = 11, \|B\|_F = 3\sqrt{3}, \|B\|_M = 4, \|B\|_1 = 5, \|B\|_2 = \sqrt{23/2 + \sqrt{429}/2}, \\ \|B\|_\infty = 6, \|C\|_S = 16, \|C\|_F = 4\sqrt{3}, \|C\|_M = 5, \|C\|_1 = 7, \|C\|_2 = \sqrt{21 + \sqrt{291}}, \\ \|C\|_\infty = 9.$$

$$10. \rho(A) = 1, \rho(B) = 15, \rho(C) = 1 + \sqrt{14}, \rho(D) = 5, \rho(F) = \sqrt{11}, \rho(G) = \sqrt{5}, \\ \rho(H) = 1, \rho(J) = 1 + \sqrt{2}, \rho(K) = 3, \rho(L) = \sqrt{13}, \rho(M) = 5.$$

$$11. \mathbf{z} = [2, -2, 7]^T.$$

$$12. \mathbf{z} = \left[\frac{3}{2}, \frac{5}{2}, 3\right]^T.$$

$$13. \mathbf{z} = \left[-\frac{34}{41}, -\frac{1}{41}, \frac{33}{41}, \frac{167}{41} \right]^T.$$

$$14. \mathbf{z} = \left[-\frac{15}{34}, \frac{9}{34}, -\frac{15}{34}, -\frac{9}{34} \right]^T.$$

$$15. \mathbf{z} = \left[-\frac{3}{5}, \frac{4}{5}, \frac{6}{5}, \frac{3}{5} \right]^T.$$

Testy ke kapitole 1

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Které čtyři matematické operace se využívají při numerickém řešení úloh?

Sčítání, odečítání, násobení a umocňování.

Sčítání, odečítání, derivování a integrování.

Sčítání, odečítání, násobení a dělení.

Logaritmování, derivování, integrování a umocňování.

2. (1b.) Co zadáváme na vstup numerické úlohy a co získáváme na výstupu?
Čísla/funkce. Čísla/čísla. Funkce/čísla. Funkce/funkce.
3. (1b.) Dokončete větu tak, aby byla pravdivá. Korektní úloha je dobře podmíněna, jestliže
je poměr absolutní chyby výstupu a vstupu malý.
je poměr čísel na výstupu a vstupu malý.
je poměr relativní chyby vstupu a výstupu malý.
je poměr relativní chyby výstupu a vstupu malý.
4. (1b.) Rozhodněte, zda je následující tvrzení pravdivé. Jestliže je algoritmus numericky stabilní, pak je málo citlivý na vliv zaokrouhlovacích chyb, které vznikají během výpočtu.
Ano. Ne.

5. (1b.) Je dán vektor $x = [1, -2, -8, 7, 4]^T \in \mathbb{R}^5$. Pak je jeho maximální norma rovna:

2.

8.

7.

 $\sqrt{134}$.

6. (1b.) Vektorová norma $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ se nazývá:

Eukleidovská.

Součtová.

Frobeniova.

Maximální.

7. (1b.) Na množině matic $\mathbb{M}_{m,n}$ lze zavést tzv. operátorové normy díky tomu, že je na ní definována operace ... matic se sloupcovými vektory délky n .

sečítání,

umocňování,

násobení,

odečítání.

8. (1b.) Vyberte tu vlastnost, kterou nesplňuje norma $\|x\|$ nenulového vektoru $x \in \mathbb{R}^n$.

$$\|c \cdot x\| = |c| \cdot \|x\|, c \in \mathbb{R}.$$

$$\|x + y\| \leq \|x\| + \|y\|, \text{ pro každý vektor } y \in \mathbb{R}^n.$$

$$\|x\| = 0.$$

$$\|x\| > 0.$$

9. (1b.) Spektrální poloměr $\rho(A)$ matice $A \in \mathbb{M}_n$ je roven hodnotě

$$\max_{i=1,\dots,n} |\lambda_i|, \text{ kde } \lambda_i \text{ jsou vlastní čísla matice } A.$$

$$\max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|, \text{ kde } a_{ij} \text{ jsou prvky matice } A.$$

$$\max_{\substack{i=1,\dots,n, \\ j=1,\dots,n}} \{|a_{ij}|\}, \text{ kde } a_{ij} \text{ jsou prvky matice } A.$$

10. (1b.) Která z vlastností platí pro skalární součin na \mathbb{R}^n ?

$$(\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{y} + \mathbf{z}). \quad (\mathbf{x}, \mathbf{x}) \leq 0.$$

$$(\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z}). \quad (c\mathbf{x}, c\mathbf{y}) = c(\mathbf{x}, \mathbf{y}), \text{ kde } c \in \mathbb{R}.$$

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (1b.) Vyberte zjednodušené schéma numerického řešení nějaké úlohy.

Reálný problém – numerická úloha – numerický algoritmus – matematický model.

Reálný problém – matematický model – numerická úloha – numerický algoritmus.

Experiment – analytické řešení – matematický model – numerický algoritmus.

Reálný problém – numerický model – matematický model – numerický algoritmus.

2. (1b.) Je-li x přesné číslo a \tilde{x} jeho přibližná hodnota, pak rozdíl $x - \tilde{x} = \Delta x$ nazýváme
absolutní chybou aproximace \tilde{x} . odhadem relativní chyby.

relativní chybou aproximace \tilde{x} . odhadem absolutní chyby.

3. (1b.) Dokončete větu tak, aby byla pravdivá. Je-li číslo podmíněnosti úlohy $C_p \approx 150$, pak je úloha
- | | |
|-------------------|----------------|
| dobře podmíněná. | chybně zadaná. |
| špatně podmíněná. | neřešitelná. |
4. (1b.) Doplňte tvrzení tak, aby bylo pravdivé. Norma libovolného vektoru $x \in \mathbb{R}^n$ je
- | | |
|---------------------|---------------------|
| větší než 0. | menší než 0. |
| menší nebo rovna 0. | větší nebo rovna 0. |
5. (1b.) Je dán vektor $x = [-1, 3, -4, -5, -2, 3]^T \in \mathbb{R}^6$. Pak je jeho eukleidovská norma rovna:
- | | | | |
|-----|----|----|-----|
| -7. | 5. | 8. | 18. |
|-----|----|----|-----|
6. (1b.) Vektorová norma $\|x\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$ se nazývá:
- | | | | |
|-----------|-------------|---------------|------------|
| Součtová. | Frobeniova. | Eukleidovská. | Maximální. |
|-----------|-------------|---------------|------------|

7. (1b.) Maticové normy se zavádějí pouze pro čtvercové matice \mathbb{M}_n .

Ano.

Ne.

8. (1b.) Vypočtete maximální normu matice

$$A = \begin{pmatrix} -6 & 2 & -1 & 0 & 3 \\ 1 & -2 & 2 & 4 & -5 \end{pmatrix}.$$

5.

6.

26.

$\sqrt{98}$.

9. (1b.) Číslo podmíněnosti matice ze základního definičního vztahu lze určit pro libovolnou matici.

čtvercovou matici.

regulární matici.

matici, jejíž determinant je roven jedné.

10. (1b.) V případě, že pro dva nenulové vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, kde \mathbb{R}^n je prostor se skalárním součinem, platí, že $(\mathbf{x}, \mathbf{y}) = 0$, pak

je jeden vektor násobkem druhého.

jsou vektory k sobě kolmé.

norma jednoho z vektorů je rovna nule.

tato skutečnost nemůže nikdy nastat.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (1b.) Jaký je vztah mezi matematickým modelem a numerickou úlohou?

Výsledkem zanedbáním některých rovnic v matematickém modelu je numerická úloha.

Jestliže v matematickém modelu zaokrouhlíme naměřené veličiny a fyzikální konstanty, získáme numerickou úlohu.

Numerická úloha je výsledkem digitalizace matematického modelu.

Numerická úloha je postup řešení matematického modelu.

2. (1b.) Označme x přesné číslo a \tilde{x} jeho přibližnou hodnotu. Vyberte vztah pro výpočet relativní chyby aproximace \tilde{x} .

$$\frac{x - \tilde{x}}{x}.$$

$$x - \tilde{x}.$$

$$\left| \frac{x - \tilde{x}}{x} \right|.$$

$$\left| \frac{\tilde{x}}{x} \right|.$$

3. (1b.) Kdy mohou vznikat při řešení numerické úlohy zaokrouhlovací chyby?

Při měření vstupních dat empirických hodnot.

Při nevhodné volbě matematického modelu.

Při provádění aritmetických operací na počítači.

Při aproximaci matematického modelu numerickou úlohou.

4. (1b.) Co znamená, že numerický algoritmus je dobře podmíněný?

Je málo citlivý na poruchy ve vstupních datech.

Je málo citlivý na vliv zaokrouhlovacích chyb vznikajících při výpočtu.

5. (1b.) Vyberte tu vlastnost, kterou vektorová norma na \mathbb{R}^n nemá (\mathbf{x} a \mathbf{y} jsou vektory z \mathbb{R}^n a $c \in \mathbb{R}$).

$$\|\mathbf{x}\| \geq 0.$$

$$\|c\mathbf{x}\| = |c| \cdot \|\mathbf{x}\|.$$

$$\|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{x}\| - \|\mathbf{y}\|.$$

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

6. (1b.) Je dán vektor $x = [-1, 3, -2, -7, 4, 6]^T \in \mathbb{R}^6$. Pak je jeho součtová norma rovna:

3.

7.

23.

 $\sqrt{87}$.

7. (1b.) Vyberte předpis, který se používá pro výpočet Frobeniovy normy ($A \in \mathbb{M}_{m,n}$, $x \in \mathbb{R}^n$).

$$\|A\|_M = \max_{\substack{i=1,\dots,m, \\ j=1,\dots,n}} \{|a_{ij}|\}.$$

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}.$$

$$\|x\|_1 = |x_1| + \dots + |x_n|.$$

$$\|x\|_2 = \sqrt{x_1^2 + \dots + x_n^2}.$$

8. (1b.) Je-li A čtvercová matice a λ_i , kde $i = 1, \dots, n$, jsou její vlastní čísla, pak číslo $\rho(A) = \max_{i=1,\dots,n} |\lambda_i|$ se nazývá

číslo podmíněnosti matice A .spektrální norma matice A .relaxační poloměr matice A .spektrální poloměr matice A .

9. (1b.) Vyberte předpis pro standardní skalární součin na \mathbb{R}^n .

$$(\mathbf{x}, \mathbf{y}) = |x_1 y_1 + \cdots + x_n y_n|. \quad (\mathbf{x}, \mathbf{y}) = \sqrt{x_1 y_1 + \cdots + x_n y_n}.$$

$$(\mathbf{x}, \mathbf{y}) = x_1 y_1 + \cdots + x_n y_n.$$

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 4

1. (1b.) Dokončete větu tak, aby byla pravdivá. Numerický algoritmus je systém lineárních nebo nelineárních rovnic, pomocí kterých vyřešíme reálný problém.

přesný popis kroků, které vedou k vyřešení numerické úlohy.

postup, pomocí kterého ze zadaných funkcí na vstupu získáme číselné hodnoty na výstupu.

postup, při kterém získáme výsledek matematické úlohy.

2. (1b.) Správně zaokrouhlená hodnota čísla 99,4253827 na tři platné cifry bude

99,425.

99,0.

100.

99,4.

3. (1b.) Je dán libovolný nenulový vektor $\mathbf{x} \in \mathbb{R}^n$. Pak jeho norma $\|\mathbf{x}\|$ může být rovna nule.

jeho norma $\|\mathbf{x}\|$ může být libovolné záporné číslo.

jeho norma $\|\mathbf{x}\|$ nemůže být záporné číslo.

o znaménku normy $\|\mathbf{x}\|$ nelze nic říct.

4. (1b.) Vyberte vektor, jehož součtová i maximální norma jsou si rovny.

$$\mathbf{x} = (0, 0, 2, 0, 0).$$

$$\mathbf{x} = (1/2, 0, 1, 0, -1/2).$$

$$\mathbf{x} = (1/3, -1/3, 1/3, -1/3, 1/3).$$

Žádná z nabízených možností není správná.

5. (1b.) Vyberte předpis, který nepředstavuje žádnou vektorovou normu vektoru \mathbf{x} nebo maticovou normu matice A (λ_i jsou vlastní čísla matice A).

$$\max\{|x_1|, \dots, |x_n|\}.$$

$$\max_{\substack{i=1, \dots, m \\ j=1, \dots, n}} |a_{ij}|.$$

$$\sqrt{\sum_{i=1}^n x_i^2}.$$

$$\max_{i=1, \dots, n} |\lambda_i|.$$

6. (1b.) Vypočtěte součtovou normu matice

$$A = \begin{pmatrix} -2 & -3 & -1 & 0 & -4 \\ 1 & -2 & 2 & 4 & -1 \end{pmatrix}.$$

-6.

4.

5.

20.

7. (1b.) Je-li matice A dobře podmíněná, znamená to, že pro její číslo podmíněnosti $\kappa(A)$ platí

$$\kappa(A) < 0.$$

$$\kappa(A) \approx 1.$$

$$\kappa(A) \gg 1.$$

$$\kappa(A) = 1.$$

8. (1b.) Spektrální poloměr $\rho(A)$ matice A lze určit pouze pro
obdélníkovou matici. čtvercovou matici.

regulární matici.

9. (1b.) Která z vlastností platí pro vektorovou normu na \mathbb{R}^n ?

$$\|\mathbf{x}\| \leq 0.$$

$$\|c\mathbf{x}\| = c\|\mathbf{x}\|, \text{ kde } c \in \mathbb{R}.$$

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

$$\|\mathbf{x} + \mathbf{y}\| + \|\mathbf{z}\| \leq \|\mathbf{x}\| + \|\mathbf{y} + \mathbf{z}\|.$$

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Kapitola 2

Numerické řešení nelineárních rovnic

Cíle

Po prostudování této kapitoly budete schopni vysvětlit:

- jak postupujeme při hledání kořenů nelineárních rovnic a jejich soustav,
- jak posuzujeme kvalitu metod a přesnost řešení,
- podle čeho metody klasifikujeme,
- které jsou nejdůležitější metody pro řešení nelineárních rovnic,
- které jsou nejdůležitější metody pro řešení soustav nelineárních rovnic.

Řešení rovnic patří k základním úlohám numerické matematiky. S potřebou najít kořeny rovnic se setkáváme v mnoha partiích matematiky, avšak jen výjimečně dokážeme jejich hodnoty přesně určit. V této kapitole se budeme věnovat řešení nelineárních rovnic, studium soustav lineárních rovnic, které má svá specifika, odložíme do kapitoly 3.

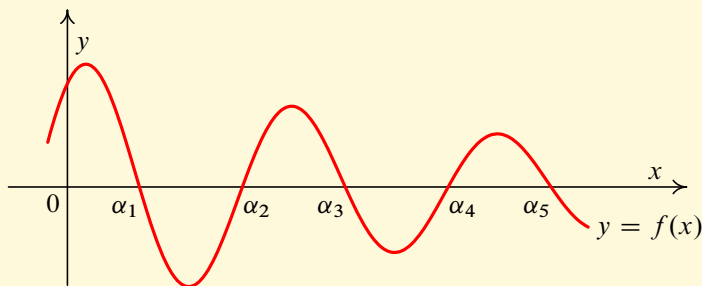
Ve většině kapitoly se budeme zabývat jednou rovnicí o jedné neznámé. Seznámíme se se základními principy a pojmy a uvedeme tradiční metody řešení takových rovnic. Závěrečný oddíl je věnován soustavám nelineárních rovnic, tato problematika je však daleko složitější a podrobnější výklad přesahuje cíle tohoto textu.

2.1 Nelineární rovnice o jedné neznámé

Budeme se zabývat hledáním reálných kořenů rovnice

$$f(x) = 0. \quad (2.1)$$

Naším úkolem je najít čísla $\alpha \in R$, pro něž platí, že $f(\alpha) = 0$. Tato čísla nazýváme *kořeny rovnice*. Chceme vlastně najít x -ové souřadnice průsečíků grafu funkce f s osou x

Obr. 2.1: Kořeny $\alpha_1, \dots, \alpha_5$ rovnice $f(x) = 0$

— viz obr. 2.1. Obecněji bychom mohli uvažovat rovnici

$$g(x) = h(x).$$

Tu lze ale vždy upravit na tvar (2.1), kde $f(x) = g(x) - h(x)$.

O funkci jedné proměnné $f(x)$ budeme předpokládat, že je „rozumná“, tj. je aspoň spojitá. Tato vlastnost nám zaručí existenci kořenů. Má-li spojitá funkce ve dvou bo-

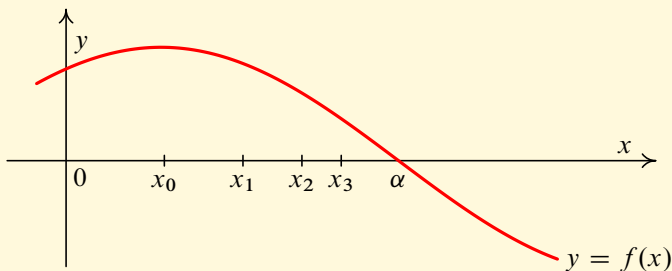
dech opačná znaménka, má podle Cauchyovy-Bolzanovy¹ věty mezi nimi kořen (viz [33, str. 204]). Bez předpokladu spojitosti v podstatě nemá úloha smysl.

2.1.1 Postup při hledání kořenů

Hledání kořenů lze v zásadě rozdělit na následující dva kroky.

1. *Separace kořenů* — určení intervalu, který je dostatečně malý a obsahuje jediný kořen.
2. *Aproximace kořenů* — nalezení posloupnosti čísel x_0, x_1, x_2, \dots , která konverguje ke kořenu α , tj. $\lim_{n \rightarrow \infty} x_n = \alpha$, $f(\alpha) = 0$, viz obr. 2.2. Sestrojení takové posloupnosti se dělá pomocí některé z iteračních metod, které budou uvedeny v následujícím textu.

¹**Bernard Placidus Johann Nepomuk Bolzano** (1781–1848) — filozof a největší český matematik 19. století. Působil jako profesor filozofie a náboženství na Karlově univerzitě. Většina jeho matematických prací zůstala v rukopisech a nebyla tehdy známa, takže neovlivnila další rozvoj matematiky. Jeho výsledky byly později znovuobjeveny. Přes tragický osud jeho díla je mu dnes přiznána priorita v mnoha myšlenkách a je po něm pojmenována řada tvrzení zejména diferenciálního počtu.



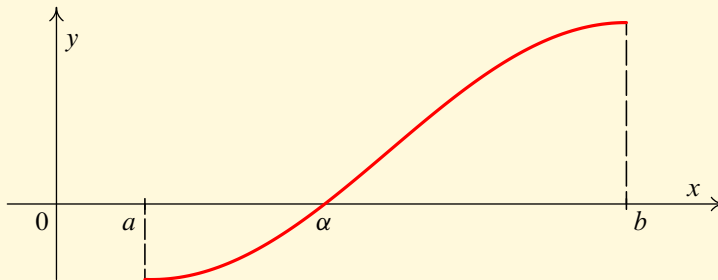
Obr. 2.2 Aproximace kořenu

Pro separaci kořenů neexistuje jednoznačná a spolehlivá univerzální metoda. Obvykle informaci o poloze kořenů získáváme některým z následujících postupů.

- Graficky — z obrázku grafu přibližně odhadneme polohu kořenů. Tento způsob je dnes, kdy jsou k dispozici počítačové programy, které graf dokážou snadno a s velkou přesností nakreslit, nejčastější.
- Pomocí tabulky hodnot — vypočítáme funkční hodnoty ve vhodných bodech x a vybereme ty intervaly, v jejichž krajních bodech mají funkční hodnoty opačná znaménka.

Tento postup byl typický v minulosti, kdy nebyly k dispozici počítače, ale pouze kalkulačky, popř. jen ruční výpočty a pomůcky, které je usnadňovaly (logaritmické pravítko, tabulky logaritmů, goniometrických funkcí apod.).

- (c) Z vlastností problému, který je uvažovanou rovnicí modelován.
- (d) Analýzou vlastností funkce nástroji diferenciálního počtu. Je-li $f(a) \cdot f(b) < 0$, $a < b$, pak mezi a a b existuje kořen. Monotonie funkce f na intervalu $\langle a, b \rangle$



Obr. 2.3: Existence a jednoznačnost kořenu

potom zaručuje jednoznačnost, viz obr. 2.3. Její ověření je obecně obtížné. (Monotonii lze sice určit pomocí znaménka derivace, k tomu jsou ale obvykle potřebné kořeny funkce $f'(x)$, které často neumíme určit jinak než numericky, čímž se dostáváme do začarovaného kruhu.)

2.1.2 Zastavovací podmínky

Při použití iterační metody sestrojíme podle nějakého vzorce posloupnost čísel $\{x_n\}$. Nejčastěji se hodnota x_n počítá z předcházející hodnoty x_{n-1} nebo dvojice předcházejících hodnot x_{n-1} a x_{n-2} , jsou ale i složitější vzorce. S konkrétními příklady se seznámíme dále.

Nicméně na počítači pochopitelně můžeme vypočítat pouze konečný počet členů posloupnosti $\{x_n\}$. V určitý okamžik proces musíme ukončit. Obvykle se zadá malé kladné číslo ε a posuzuje se, zda velikost nějaké veličiny je již menší než toto číslo. Ideální by bylo posuzovat vzdálenost $|x_n - \alpha|$ od přesné hodnoty kořenu α , ten však neznáme. V praxi proto k zastavení iteračního procesu používáme některou z následujících *zastavovacích podmínek* (je možné použít i jejich kombinaci):

$$\text{i) } |x_n - x_{n-1}| < \varepsilon,$$

$$\text{ii) } \frac{|x_n - x_{n-1}|}{|x_n|} < \varepsilon,$$

$$\text{iii) } |f(x_n)| < \varepsilon.$$

Jakmile je pro některé n zvolená podmínka splněna, proces ukončíme a klademe $\alpha \doteq x_n$. Číslo ε se nazývá *přesnost řešení*.

Poznámka 2.1

1. Splnění kterékoli ze zastavovacích podmínek neznamená, že absolutní chyba řešení $|\alpha - x_n|$ je menší než zvolené ε . Přesný kořen α může být ještě hodně vzdálen. Např. je-li graf funkce f v okolí kořenu plochý, může být hodnota $|f(x_n)|$ velmi malá, takže při použití podmínky iii) je proces zastaven, přestože kořen je ještě příliš daleko. Podobně je tomu u druhých dvou zastavovacích podmínek. Pro některé iterační metody existují odhady absolutní chyby $|\alpha - x_n|$ pomocí čísla ε , bývají však často pesimistické (skutečná chyba je daleko menší než odhad).
2. Pokud je kořen α blízký nule nebo je naopak hodně velký, je vhodné použít zastavovací podmínku ii).

3. Při velmi pomalé konvergenci je vhodné nastavit maximální povolený počet iterací s následnou signalizací, že nebyla dodržena požadovaná přesnost. Tím se vyhneme možnému nekonečnému cyklu (v důsledku zaokrouhlovacích chyb se může stát, že požadovaná zastavovací podmínka nebude nikdy splněna).

2.1.3 Rychlost konvergence

Sestrojíme-li pomocí nějaké iterační metody posloupnost $\{x_n\}$, zajímají nás odpovědi na následující otázky:

1. Konverguje tato posloupnost ke kořenu α rovnice $f(x) = 0$?
2. Jestliže ano, jak rychle?

Podmínkami konvergence se budeme zabývat u jednotlivých metod. Nyní si všimneme toho, jak se posuzuje rychlost, s níž se posloupnost $\{x_n\}$ přibližuje ke kořenu α . Jde o důležitý ukazatel kvality iterační metody.

Předpokládejme, že posloupnost x_0, x_1, x_2, \dots konverguje ke kořenu α . Jestliže

existuje konstanta $C > 0$ a číslo k takové, že

$$\lim_{n \rightarrow \infty} \frac{|x_{n+1} - \alpha|}{|x_n - \alpha|^k} = C,$$

nazývá se k řád konvergence posloupnosti a C chybová konstanta.

Říkáme, že daná iterační metoda je řádu k , jestliže libovolná konvergentní posloupnost získaná touto metodou má řád konvergence větší nebo rovný k a alespoň jedna posloupnost má řád konvergence přesně k .

Je-li řád konvergence posloupnosti $\{x_n\}$ roven k , platí v blízkosti kořenu, tj. pro velká n , přibližný vztah $|x_{n+1} - \alpha| \approx C|x_n - \alpha|^k$. Protože pro velká n je $|x_n - \alpha|$ malé číslo, je pro velké k číslo $|x_n - \alpha|^k$ ještě mnohem menší, tudíž absolutní chyba $|x_{n+1} - \alpha|$ bude daleko menší než absolutní chyba $|x_n - \alpha|$. Tedy čím vyšší je řád konvergence, tím rychleji se x_n přibližuje k α .

Používá se terminologie *lineární* ($k = 1$, $C < 1$), *superlineární* ($k > 1$) a *kvadratická* ($k = 2$) konvergence.

2.2 Iterační metody pro řešení rovnic o jedné neznámé

V tomto oddílu uvedeme pět klasických iteračních metod pro řešení nelineárních rovnic o jedné neznámé. Tento výčet zdaleka není úplný, v literatuře lze nalézt řadu dalších metod, viz např. [22, 43, 52].

2.2.1 Klasifikace iteračních metod

Iterační metody se obvykle dělí na dva typy:

- i) *Startovací metody* — jsou vždy konvergentní, ale konvergence je pomalá.
- ii) *Zpřesňující metody* — jsou rychlejší, ale počáteční aproximace musí být dostatečně blízko kořenu, aby metoda konvergovala.

Obvykle postupujeme tak, že provedeme separaci kořenu, tj. určíme dostatečně malý interval obsahující jediný kořen dané rovnice. Pak některou startovací metodou najdeme přibližnou hodnotu tohoto kořenu. Tu pak zlepšíme vhodnou zpřesňující metodou.

Z pěti níže uváděných metod se mezi startovací řadí metoda bisekce (metoda půlení intervalu) a metoda regula falsi (metoda tětív), které konvergují za dosti obecných předpo-

kladů, ale jejich konvergence bývá pomalá. Proto se používají pouze k hrubé aproximaci kořenu.

Mezi zpřesňující metody patří metoda prosté iterace, Newtonova metoda (metoda tečen) a metoda sečen.

V konkrétním případě může být startovací metoda rychlá a zpřesňující metoda pomalá. Klasifikace není jednoznačná, někdy se metoda prosté iterace řadí mezi startovací metody. Navíc tuto metodu lze chápat jako jakési obecné východisko pro konstrukci dalších metod. Ani vyšší řád konvergence nemusí u konkrétní rovnice nutně znamenat rychlejší konvergenci.

U každé metody kromě jejího popisu a podmínek konvergence uvedeme ukázkový příklad. Jeho účelem je především pochopit, jaký je geometrický princip dané metody. K tomu slouží zařazené animace. Aby bylo dobře vidět jednotlivé kroky, byly úmyslně vybrány poněkud „neobvyklé“ rovnice a počáteční aproximace byly voleny ne příliš blízko hledaného kořenu. Jinak by totiž další aproximace byly rychle příliš dobré a v animacích bychom toho moc neviděli. Výpočty byly provedeny v programu Maple 2018, počet dekadických cifer mantisy byl nastaven na 10.

2.2.2 Metoda bisekce (metoda půlení intervalu)

Předpokládáme, že funkce $f(x)$ je spojitá na intervalu $\langle a, b \rangle$ a platí $f(a)f(b) < 0$. To zajišťuje existenci aspoň jednoho kořenu rovnice $f(x) = 0$ uvnitř tohoto intervalu. Následující konstrukce nevyžaduje, aby kořen byl jediný.

Postup:

Označme $a_0 = a$, $b_0 = b$. Sestrojíme posloupnost vnořených intervalů, které vždy obsahují kořen rovnice, přičemž jejich délka se zmenšuje a konverguje k nule.

- (1) Určíme střed $c_0 = \frac{a_0 + b_0}{2}$ intervalu $\langle a_0, b_0 \rangle$.

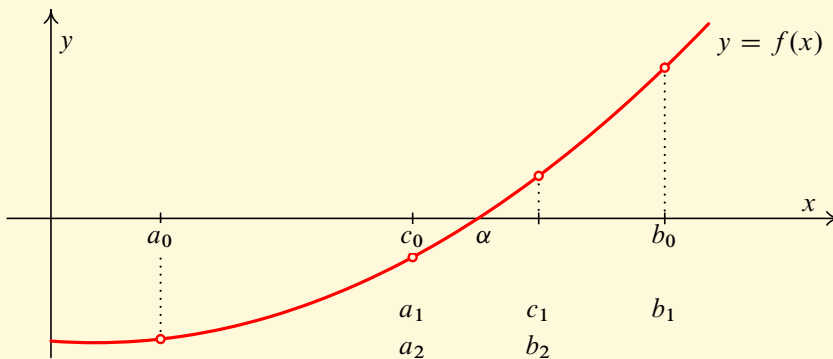
Je-li $f(c_0) = 0$, našli jsme kořen $\alpha = c_0$ a postup končí.

V opačném případě některý z intervalů $\langle a_0, c_0 \rangle$, $\langle c_0, b_0 \rangle$ obsahuje kořen. Platí-li $f(a_0)f(c_0) < 0$, je to interval $\langle a_0, c_0 \rangle$, je-li $f(c_0)f(b_0) < 0$, je to interval $\langle c_0, b_0 \rangle$. Tento interval označíme $\langle a_1, b_1 \rangle$. Zřejmě platí $b_1 - a_1 = \frac{b_0 - a_0}{2}$ a $f(a_1)f(b_1) < 0$.

- (2) Postup opakujeme. Obecně v n -tém kroku vycházíme z intervalu $\langle a_{n-1}, b_{n-1} \rangle$, pro nějž platí $b_{n-1} - a_{n-1} = \frac{b_0 - a_0}{2^{n-1}}$ a $f(a_{n-1})f(b_{n-1}) < 0$.

Určíme střed $c_{n-1} = \frac{a_{n-1} + b_{n-1}}{2}$ intervalu $\langle a_{n-1}, b_{n-1} \rangle$.

Je-li $f(c_{n-1}) = 0$, našli jsme kořen $\alpha = c_{n-1}$ a postup končí.



Obr. 2.4: Metoda bisekce

V opačném případě některý z intervalů $\langle a_{n-1}, c_{n-1} \rangle$, $\langle c_{n-1}, b_{n-1} \rangle$ obsahuje kořen. Jestliže platí $f(a_{n-1})f(c_{n-1}) < 0$, je to interval $\langle a_{n-1}, c_{n-1} \rangle$, jestliže platí $f(c_{n-1})f(b_{n-1}) < 0$, je to interval $\langle c_{n-1}, b_{n-1} \rangle$. Tento interval označíme $\langle a_n, b_n \rangle$. Zřejmě platí $b_n - a_n = \frac{b_{n-1} - a_{n-1}}{2} = \frac{b_0 - a_0}{2^n}$ a $f(a_n)f(b_n) < 0$.

(3) Na obr. 2.4 jsou znázorněny první dva kroky.

Teoreticky můžeme po konečném počtu kroků najít přesnou hodnotu kořenu, kromě velmi jednoduchých rovnic je to však vzhledem k zaokrouhlovacím chybám nepravděpodobné.

Za členy posloupnosti, která konverguje ke kořenu α , zvolíme středy intervalů $\langle a_n, b_n \rangle$, tj. $x_n = \frac{a_n + b_n}{2}$, $n = 0, 1, 2, \dots$

Konvergence:

Jedná se o vždy konvergentní metodu. Zřejmě $a_0 \leq a_1 \leq a_2 \leq \dots$ (posloupnost je neklesající), $b_0 \geq b_1 \geq b_2 \geq \dots$ (posloupnost je nerostoucí) a $a_n < b_n$. Obě posloupnosti $\{a_n\}$ a $\{b_n\}$ jsou proto konvergentní. Protože $\lim_{n \rightarrow \infty} (b_n - a_n) = \lim_{n \rightarrow \infty} \frac{b_0 - a_0}{2^n} = 0$, musí mít stejnou limitu. Jelikož libovolný interval $\langle a_n, b_n \rangle$ obsahuje kořen rovnice $f(x) = 0$, musí platit $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = \alpha$, kde α je kořen rovnice.

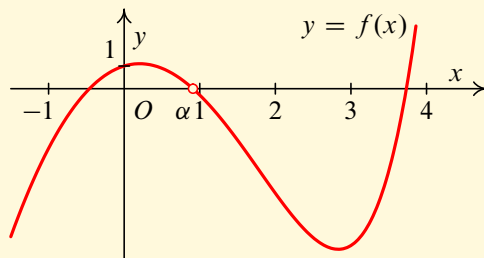
Protože $a_n < x_n < b_n$, platí rovněž $\lim_{n \rightarrow \infty} x_n = \alpha$. Metoda bisekce tedy vždy konverguje. Konvergence je obecně pomalá, platí $|x_n - \alpha| \leq \frac{b_0 - a_0}{2^{n+1}}$. Nevyužívají se žádné speciální vlastnosti funkce f , pouze skutečnost, zda funkční hodnoty jsou kladné nebo záporné. Máme však dolní i horní odhad pro velikost kořenu, protože $a_n < x_n < b_n$.

Pokud výchozí interval $\langle a, b \rangle$ obsahoval více kořenů, metoda dává posloupnost $\{x_n\}$, která sice konverguje ke kořenu rovnice $f(x) = 0$, nevíme ovšem, ke kterému. Když výpočet po konečném počtu kroků zastavíme, může být v tomto případě ve výsledném intervalu $\langle a_n, b_n \rangle$ více kořenů.

Příklad 2.2 Metodou bisekce najděte nejmenší kladný kořen rovnice $e^x - 3x^2 = 0$. Použijte zastavovací podmínku $|f(x_n)| < \varepsilon$, kde $\varepsilon = 0,02$.

Řešení. Nejprve nakreslíme graf funkce $f(x) = e^x - 3x^2$ — viz obr. 2.5. Z obrázku je vidět, že hodnota nejmenšího kladného kořenu α je $\alpha \approx 1$. Zvolíme tedy $a_0 = 0,5$ a $b_0 = 2$. V intervalu $\langle a_0, b_0 \rangle$ má rovnice jediný kořen.

Nyní podle výše popsaného postupu určíme intervaly $\langle a_1, b_1 \rangle$, $\langle a_2, b_2 \rangle$ atd. Výsledky jsou zapsány v následující tabulce. Je vidět, že zastavující podmínka byla splněna po sedmi krocích. Platí $x_6 = 0,96875$, tedy $\alpha \approx 0,96875$.



Obr. 2.5: Graf funkce z příkladu 2.2

n	a_n	b_n	x_n	$f(x_n)$	$ f(x_n) $
0	0,5000000000	2,0000000000	1,2500000000	-1,197157043	1,197157043
1	0,5000000000	1,2500000000	0,8750000000	0,102000294	0,102000294
2	0,8750000000	1,2500000000	1,0625000000	-0,493122806	0,493122806
3	0,8750000000	1,0625000000	0,9687500000	-0,180780599	0,180780599
4	0,8750000000	0,9687500000	0,9218750000	-0,035560824	0,035560824
5	0,8750000000	0,9218750000	0,8984375000	0,034193158	0,034193158
6	0,8984375000	0,9218750000	0,9101562500	-0,000442459	0,000442459

U metody bisekce můžeme odhadnout vzdálenost získaného přibližného výsledku od přesné hodnoty kořenu. Platí totiž $|\alpha - x_6| \leq \frac{b_0 - a_0}{2^7} = \frac{1,5}{128} \doteq 0,012$.

Na obr. 2.6 je uvedena animace, která znázorňuje konstrukci jednotlivých členů posloupnosti $\{x_n\}$, jež konverguje ke kořenu α . ▲

Obr. 2.6: Animace
k příkladu 2.2 —
metoda bisekce

2.2.3 Metoda regula falsi (metoda tětív)

Předpokládáme, že funkce $f(x)$ je spojitá na intervalu $\langle a, b \rangle$ a platí $f(a)f(b) < 0$. To zajišťuje existenci aspoň jednoho kořenu rovnice $f(x) = 0$ uvnitř tohoto intervalu. Tětivou rozumíme úsečku (část přímky) spojující dva body na grafu funkce. Následující konstrukce nevyžaduje, aby kořen byl jediný.

Postup:

Označme $a_0 = a$, $b_0 = b$. Sestrojíme posloupnost vnořených intervalů, které vždy obsahují kořen rovnice, přičemž na těchto intervalech budeme graf funkce nahrazovat přímkami (tětivami) určenými hodnotami funkce f v krajních bodech intervalů.

- (1) Graf funkce f nahradíme na intervalu $\langle a_0, b_0 \rangle$ přímkou, jež prochází body $[a_0, f(a_0)]$ a $[b_0, f(b_0)]$. Lineární funkce l , jejímž grafem je tato přímka, má vzorec $l(x) = f(a_0) + k(x - a_0)$, kde $k = \frac{f(b_0) - f(a_0)}{b_0 - a_0}$ je směrnice. Protože $l(a_0)l(b_0) = f(a_0)f(b_0) < 0$, má rovnice $l(x) = 0$ v tomto intervalu jediný kořen, který označíme c_0 . Z rovnice

$$f(a_0) + \frac{f(b_0) - f(a_0)}{b_0 - a_0} (c_0 - a_0) = 0$$

vypočítáme, že

$$c_0 = a_0 - \frac{f(a_0)}{f(b_0) - f(a_0)} (b_0 - a_0). \quad (2.2)$$

Jmenovatel $f(b_0) - f(a_0)$ v předchozím vztahu je nenulový, protože čísla $f(a_0)$ a $f(b_0)$ mají opačná znaménka.

Je-li $f(c_0) = 0$, našli jsme kořen $\alpha = c_0$ a postup končí.

V opačném případě některý z intervalů $\langle a_0, c_0 \rangle$, $\langle c_0, b_0 \rangle$ obsahuje kořen. Platí-li $f(a_0)f(c_0) < 0$, je to interval $\langle a_0, c_0 \rangle$, je-li $f(c_0)f(b_0) < 0$, je to interval $\langle c_0, b_0 \rangle$. Tento interval označíme $\langle a_1, b_1 \rangle$. Zřejmě platí $f(a_1)f(b_1) < 0$.

- (2) Postup opakujeme. Obecně v n -tém kroku vycházíme z intervalu $\langle a_{n-1}, b_{n-1} \rangle$, pro nějž platí $f(a_{n-1})f(b_{n-1}) < 0$. Graf funkce f nahradíme na tomto intervalu přímkou, která prochází body $[a_{n-1}, f(a_{n-1})]$ a $[b_{n-1}, f(b_{n-1})]$. Lineární funkce l , jejímž grafem je tato přímka, má vzorec $l(x) = f(a_{n-1}) + k(x - a_{n-1})$, kde $k = \frac{f(b_{n-1}) - f(a_{n-1})}{b_{n-1} - a_{n-1}}$ je směrnice. Protože $l(a_{n-1})l(b_{n-1}) = f(a_{n-1})f(b_{n-1}) < 0$, má rovnice $l(x) = 0$ v tomto intervalu jediný kořen, který označíme c_{n-1} . Z rovnice

$$f(a_{n-1}) + \frac{f(b_{n-1}) - f(a_{n-1})}{b_{n-1} - a_{n-1}} (c_{n-1} - a_{n-1}) = 0,$$

vypočítáme, že

$$c_{n-1} = a_{n-1} - \frac{f(a_{n-1})}{f(b_{n-1}) - f(a_{n-1})} (b_{n-1} - a_{n-1}). \quad (2.3)$$

Jmenovatel $f(b_{n-1}) - f(a_{n-1})$ v předchozím vztahu je nenulový, protože čísla $f(a_{n-1})$ a $f(b_{n-1})$ mají opačná znaménka.

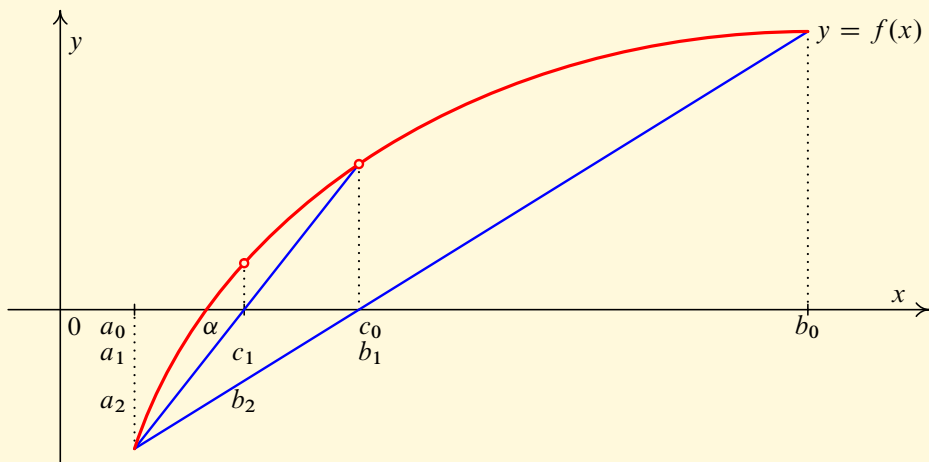
Je-li $f(c_{n-1}) = 0$, našli jsme kořen $\alpha = c_{n-1}$ a postup končí.

V opačném případě některý z intervalů $\langle a_{n-1}, c_{n-1} \rangle$, $\langle c_{n-1}, b_{n-1} \rangle$ obsahuje kořen. Jestliže platí $f(a_{n-1})f(c_{n-1}) < 0$, je to interval $\langle a_{n-1}, c_{n-1} \rangle$, jestliže platí $f(c_{n-1})f(b_{n-1}) < 0$, je to interval $\langle c_{n-1}, b_{n-1} \rangle$. Tento interval označíme $\langle a_n, b_n \rangle$. Zřejmě platí $f(a_n)f(b_n) < 0$.

(3) Na obr. 2.7 jsou znázorněny první dva kroky.

Získali jsme posloupnost vnořených intervalů $\langle a_n, b_n \rangle$, jejichž délky se zkracují. Na rozdíl od metody bisekce však tentokrát nemusí platit, že $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$. Naopak, jak uvidíme dále, v „rozumných“ případech tomu tak není.

Nyní položíme: $x_0 = a_0$, $x_1 = b_0$ (nebo naopak, pořadí není podstatné) a $x_n = c_{n-2}$ pro $n \geq 2$. Členy posloupnosti $\{x_n\}$ jsou tedy $a_0, b_0, c_0, c_1, c_2, \dots$



Obr. 2.7: Metoda regula falsi

Stručně shrnuto: Nový interval $\langle a_n, b_n \rangle$ vzniká z předchozího tak, že jeden koncový bod se zachová (zůstává pevný), druhý se změní (posune se „dovnitř“). Zda se mění

levý nebo pravý koncový bod, je obecně zcela nepravidelné (závisí to na hodnotách funkce f), je např. možné, že jeden koncový bod je pořád pevný a mění se pouze druhý. Ty koncové body, které se mění, vytvářejí posloupnost $\{c_n\}$, a tudíž i $\{x_n\}$.

- (4) Najdeme rekurentní vztah pro členy posloupnosti $\{x_n\}$. Předpokládáme, že přesnou hodnotu kořenu nenajdeme po konečném počtu kroků, tj. platí $f(x_n) \neq 0$ pro $n = 0, 1, 2, \dots$. Z (2.3) je vidět, že pro $n \geq 2$ se c_n počítá z a_n a b_n . Přitom jedno z těchto čísel je c_{n-1} . Druhé je koncovým bodem některého z předchozích intervalů $\langle a_k, b_k \rangle$, $k < n$. Při našem označení lze závislost popsat takto:

Pro $n \geq 1$ označme $m = m(n)$ (zápisem $m = m(n)$ zdůrazňujeme, že m závisí na n) největší index $m < n$ takový, že $f(x_n)f(x_m) < 0$. Na intervalu s koncovými body x_n a x_m nahradíme graf funkce f přímkou procházející body $[x_n, f(x_n)]$ a $[x_m, f(x_m)]$. Lineární funkce l , jejímž grafem je tato přímka, má vzorec $l(x) = f(x_n) + k(x - x_n)$, kde $k = \frac{f(x_n) - f(x_m)}{x_n - x_m}$ je směrnice. Protože $l(x_n)l(x_m) = f(x_n)f(x_m) < 0$, má rovnice $l(x) = 0$ mezi x_n a x_m jediný kořen x_{n+1} . Z rovnice

$$f(x_n) + \frac{f(x_n) - f(x_m)}{x_n - x_m} (x_{n+1} - x_n) = 0.$$

vypočítáme, že

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_m)} (x_n - x_m), \quad n = 1, 2, \dots, \quad (2.4)$$

neboli po úpravě

$$x_{n+1} = \frac{x_m f(x_n) - x_n f(x_m)}{f(x_n) - f(x_m)}, \quad n = 1, 2, \dots \quad (2.5)$$

Jmenovatel $f(x_n) - f(x_m)$ v předchozím vztahu je nenulový, protože čísla $f(x_n)$ a $f(x_m)$ mají opačná znaménka.

Nový člen x_{n+1} posloupnosti $\{x_n\}$ se tudíž počítá z předchozího členu x_n a ještě jednoho dřívějšího členu x_m . Metoda regula falsi je proto *dvoukroková*. Protože index m je proměnný (závisí na n a pochopitelně také na funkci f), říkáme, že jde o *nestacionární dvoukrokovou iterační metodu*.

Konvergence:

Jedná se o vždy konvergentní metodu (viz [8, str. 275], důkaz není triviální). Platí $\lim_{n \rightarrow \infty} x_n = \alpha$, kde α je kořen rovnice. Pokud výchozí interval $\langle a, b \rangle$ obsahoval více kořenů, metoda

dává posloupnost $\{x_n\}$, která sice konverguje ke kořenu rovnice $f(x) = 0$, nevíme ovšem ke kterému. Když výpočet po konečném počtu kroků zastavíme, může být v tomto případě ve výsledném intervalu $\langle a_n, b_n \rangle$ více kořenů.

Metoda regula falsi konverguje poměrně pomalu, lze ukázat, že (za předpokladu konstantního znaménka druhé derivace f'' v okolí kořenu) řád konvergence je jedna, viz [52, str. 340].

Všimneme si podrobněji, jak vypadá posloupnost $\{x_n\}$. Funkce f nabývá podle předpokladu v koncových bodech intervalu $\langle a, b \rangle$ hodnoty opačných znamének. Předpokládejme pro určitost, že $f(a) < 0$ a $f(b) > 0$. Příklad opačných znamének se vyšetří obdobně. Z výše popsané konstrukce plyne, že potom $f(a_n) < 0$ a $f(b_n) > 0$ pro libovolné $n = 0, 1, 2, \dots$. Protože $x_0 = a_0$, $x_1 = b_0$ a $x_n = c_{n-2}$ pro $n \geq 2$, je každý člen x_n , $n \geq 2$, roven některému a_k nebo b_k , kde $k \leq n - 2$.

Členy posloupnosti $\{x_n\}$ rozdělíme do dvou podposloupností. V první budou ta x_n , pro něž $f(x_n) < 0$, ve druhé ta, pro něž $f(x_n) > 0$. Ani jedna podposloupnost nebude prázdná, protože $f(x_0) < 0$ a $f(x_1) > 0$. Členy první podposloupnosti označíme x_{r_k} a druhé x_{s_k} . Zřejmě $r_0 = 0$ a $s_0 = 1$. Protože intervaly $\langle a_n, b_n \rangle$ jsou do sebe vnořené, bude první podposloupnost neklesající

a druhá nerostoucí, tj.

$$x_0 \geq x_{r_1} \geq x_{r_2} \geq x_{r_3} \geq \cdots \geq x_{s_3} \geq x_{s_2} \geq x_{s_1} \geq x_1. \quad (2.6)$$

Mohou nastat dva případy:

- (i) Jedna podposloupnost je konečná a druhá nekonečná. Nechť je nekonečná např. podposloupnost $\{x_{r_k}\}$.

K tomu dojde, právě když existuje číslo p takové, že index $m(n)$ ze vztahu (2.4) bude počínaje hodnotou p konstantní, tj. $m(p) = m(n)$ pro $n \geq p$. (Jinými slovy, pro velká k se pravé konce intervalů $\langle a_k, b_k \rangle$ přestanou měnit a členy x_n budou rovny levým koncům těchto intervalů.)

Podposloupnost $\{x_{r_k}\}$ je neklesající a ohraničená, takže má limitu. Nechť $\lim_{k \rightarrow \infty} x_{r_k} = \alpha$.

Limitním přechodem ve vztahu (2.4) dostaneme vzhledem ke spojitosti funkce f , že

$$\alpha = \alpha - \frac{f(\alpha)}{f(\alpha) - f(x_{m(p)})} (\alpha - x_{m(p)}).$$

Protože $f(x_{r_k}) < 0$, musí být $f(\alpha) \leq 0$. Dále platí $f(x_{m(p)}) > 0$. Odtud dostaneme, že jmenovatel $f(\alpha) - f(x_{m(p)})$ předchozího výrazu je nenulový a $\alpha \neq x_{m(p)}$. Z toho plyne, že nutně $f(\alpha) = 0$.

V tomto případě neplatí, že $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$.

(ii) Obě podposloupnosti jsou nekonečné.

Protože jsou obě monotónní, mají limitu. Nechť $\lim_{k \rightarrow \infty} x_{r_k} = \alpha$ a $\lim_{k \rightarrow \infty} x_{s_k} = \beta$. Vzhledem

k (2.6) musí platit $\alpha \leq \beta$, $f(\alpha) \leq 0$ a $f(\beta) \geq 0$.

Je-li $\alpha = \beta$, platí, že $\lim_{n \rightarrow \infty} x_n = \alpha$ a $f(\alpha) = 0$.

Lze ukázat, že případ $\alpha < \beta$ nemůže nastat, viz výše zmíněný odkaz [8, str. 275].

V tomto případě platí, že $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$.

Uvažujme speciální případ, kdy funkce f je na intervalu $\langle a, b \rangle$ konvexní nebo konkávní. Tuto vlastnost lze např. zjistit pomocí druhé derivace. Platí-li na $\langle a, b \rangle$ nerovnost $f''(x) \geq 0$, je zde funkce f konvexní, platí-li $f''(x) \leq 0$, je konkávní; viz [16, str. 124] nebo [33, str. 244].

Snadno lze ukázat, že jestliže funkce f je na intervalu $\langle a, b \rangle$ konvexní nebo konkávní a platí $f(a)f(b) < 0$ (tzv. *Fourierovy*¹ *podmínky*), má rovnice $f(x) = 0$ na tomto intervalu jediný kořen α , jeden z koncových bodů a_n, b_n se nemění a druhý konverguje monotónně ke kořenu α . V případě konvexní funkce se nemění ten koncový

¹**Jean-Baptiste Joseph Fourier** (1768–1830) (čti furje) — francouzský matematik, jeden ze zakladatelů matematické fyziky.

bod $d \in \{a, b\}$, v němž $f(d) > 0$, zatímco u konkávní funkce je to ten koncový bod d , v němž $f(d) < 0$. Má-li funkce f nenulovou druhou derivaci a d je ten koncový bod, pro nějž $f(d)f''(d) > 0$, pak se tento koncový bod nemění.

Funkce na obr. 2.7 je konkávní, tj. $f''(x)$ je záporná, tudíž $f(a)f''(a) > 0$. Je pěkně vidět, že posloupnost $\{c_n\}$ je ryze monotónní (klesající) a konverguje ke kořenu α zprava, levé konce a_n jsou pořád rovny a .

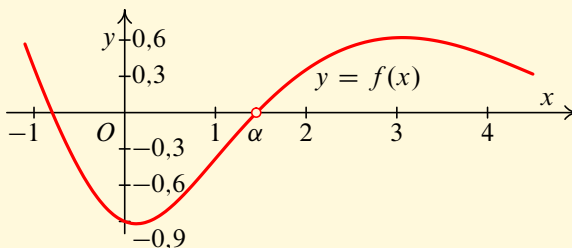
Příklad 2.3 Metodou regula falsi najděte nejmenší kladný kořen rovnice

$$\ln(x^2 + 1) - x e^{-0,2x} - 0,9 + \sin(0,7x) = 0.$$

Použijte zastavovací podmínku $\frac{|x_n - x_{n-1}|}{|x_n|} < \varepsilon$, kde $\varepsilon = 0,006$.

Řešení. Nejprve nakreslíme graf funkce $f(x) = \ln(x^2 + 1) - x e^{-0,2x} - 0,9 + \sin(0,7x)$ — viz obr. 2.8. Z obrázku je vidět, že hodnota nejmenšího kladného kořenu α je $\alpha \approx 1,5$. Zvolíme tedy $a_0 = 0,5$ a $b_0 = 4,5$. V intervalu $\langle a_0, b_0 \rangle$ má rovnice jediný kořen.

Nyní podle výše popsaného postupu s použitím vztahu (2.5) určíme intervaly $\langle a_1, b_1 \rangle$, $\langle a_2, b_2 \rangle$ atd. Výsledky jsou zapsány v následující tabulce. Je vidět, že zastavující podmínka byla splněna po pěti krocích. Platí $x_6 = 1,454\,836\,146$, tedy $\alpha \approx 1,454\,836\,146$.



Obr. 2.8: Graf funkce z příkladu 2.3

n	a_n	b_n	x_{n+2}	chyba
0	0,5	4,500000000	3,347224152	0,3443975652
1	0,5	3,347224152	2,109958583	0,5863932965
2	0,5	2,109958583	1,560007057	0,3525314347
3	0,5	1,560007057	1,462935229	0,06635415299
4	0,5	1,462935229	1,454836146	0,005567006994

Na obr. 2.9 je uvedena animace, která znázorňuje konstrukci jednotlivých členů posloupnosti $\{x_n\}$, jež konverguje ke kořenu α . ▲

Obr. 2.9: Animace
k příkladu 2.3 —
metoda
regula falsi
(metoda třetiv)

2.2.4 Metoda prosté iterace

Rovnici $f(x) = 0$, která má kořen α , upravíme na ekvivalentní (alespoň v okolí kořenu) tvar $x = g(x)$. Místo rovnice $f(x) = 0$ pak budeme řešit rovnicí $x = g(x)$. Tato rovnice bude mít rovněž kořen α , tj. musí platit $\alpha = g(\alpha)$. Hledáme tedy číslo, které po dosazení do funkce $g(x)$ dá touž hodnotu. Takovému číslu říkáme *pevný bod* funkce $g(x)$. Z geometrického hlediska hledáme průsečík grafu funkce $g(x)$ s přímkou o rovnici $y = x$.

Postup:

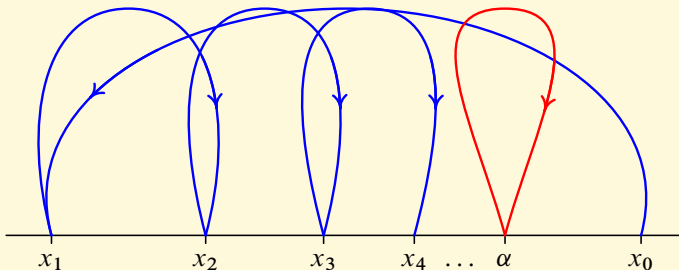
Rovnici $f(x) = 0$ upravíme na tvar $x = g(x)$ (na tzv. *tvar vhodný k iteraci*).

- (1) Zvolíme počáteční aproximaci x_0 (dostatečně blízko kořenu α). Dále položíme: $x_1 = g(x_0)$, $x_2 = g(x_1)$ atd. Obecně se další členy počítají podle vztahu

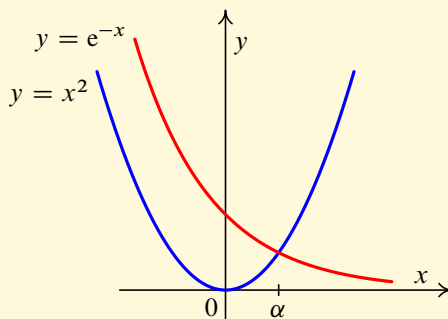
$$x_{n+1} = g(x_n) \quad n = 0, 1, 2, \dots \quad (2.7)$$

Jedná se o tzv. *stacionární jedнокrokovou iterační metodu*. Nový člen posloupnosti $\{x_n\}$ se počítá z bezprostředně předcházejícího. Pochopitelně je třeba zajistit, aby hodnota $g(x_n)$ byla vždy definovaná.

- (2) Pokud narazíme na kořen, bude pro některé k platit $x_k = g(x_k)$.
 Obecně se tak asi nestane a dostaneme posloupnost x_0, x_1, x_2, \dots . Je třeba zajistit, aby tato posloupnost konvergovala ke kořenu α .
- (3) Znázorníme-li číslo x i jeho obraz $g(x)$ na číselné ose a přiřazení vyznačíme šipkou, hledáme bod α , který se zobrazí sám na sebe. Odtud pochází název pevný bod; viz obr. 2.10, který ilustruje vytváření členů posloupnosti $\{x_n\}$, jež konverguje k pevnému bodu α .



Obr. 2.10: Princip pevného bodu

Obr. 2.11: Kořen rovnice $x^2 = e^{-x}$

Danou rovnici lze upravit mnoha způsoby. Uvažujme např. rovnici $x^2 - e^{-x} = 0$, tj. $f(x) = x^2 - e^{-x}$. Rovnici lze psát ve tvaru $x^2 = e^{-x}$. Z obr. 2.11 je zřejmé, že grafy funkcí x^2 a e^{-x} mají jediný průsečík. Uvažovaná rovnice má tudíž jediný kořen α a ten je kladný. Při převodu na tvar $x = g(x)$ se nabízejí následující formální úpravy:

$$\begin{array}{llll}
 x^2 = e^{-x} & \Rightarrow & x = \sqrt{e^{-x}} & \text{(odmocnění)} \quad g(x) = \sqrt{e^{-x}}, \\
 x^2 = e^{-x} & \Rightarrow & x = -\sqrt{e^{-x}} & \text{(odmocnění)} \quad g(x) = -\sqrt{e^{-x}}, \\
 x^2 = e^{-x} & \Rightarrow & x = -2 \ln x & \text{(logaritmování)} \quad g(x) = -2 \ln x, \\
 x^2 = e^{-x} & \Rightarrow & x = -2 \ln(-x) & \text{(logaritmování)} \quad g(x) = -2 \ln(-x), \\
 x^2 - e^{-x} = 0 & \Rightarrow & x = x + x^2 - e^{-x} & g(x) = x + f(x), \\
 e^{-x} - x^2 = 0 & \Rightarrow & x = x + e^{-x} - x^2 & g(x) = x - f(x).
 \end{array}$$

Vzhledem k tomu, že kořen α je kladný, jsou některé úpravy nesmyslné.

Konvergence:

Metoda nemusí konvergovat. Konvergence je např. zaručena, když jsou splněny předpoklady *Banachovy*¹ věty o kontrakci. Tak tomu bude, když na nějakém uzavřeném ohraničeném intervalu I obsahujícím kořen α bude existovat spojitá derivace $g'(x)$ a pro $x \in I$

¹**Stefan Banach** (1892–1945) — významný polský matematik, jeden ze zakladatelů funkcionální analýzy.

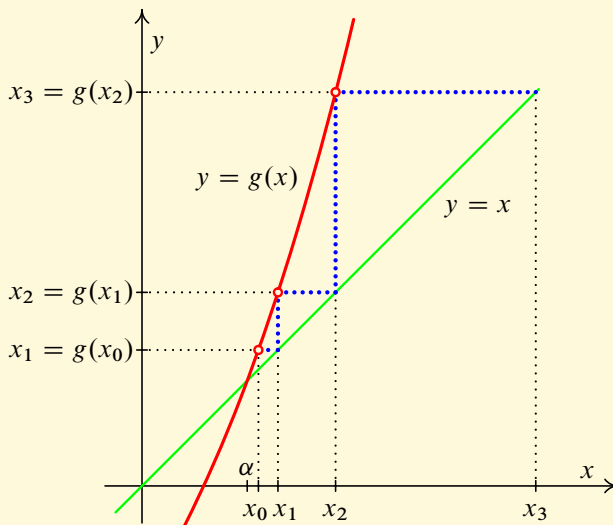
bude platit $g(x) \in I$ (tj. interval se bude zobrazovat do sebe) a $|g'(x)| < 1$ (viz [22, str. 28]).

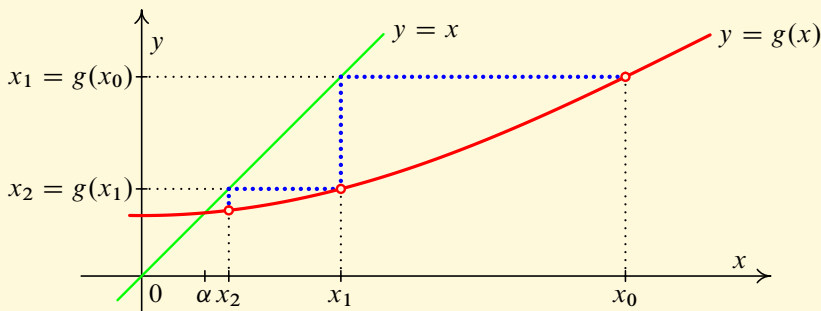
Metoda je obecně pouze řádu jedna, ale při vhodné volbě funkce $g(x)$ může být řád daleko větší. Čím menší je $|g'(x)|$ v okolí kořenu, tím je konvergence rychlejší. Pokud je absolutní hodnota derivace v okolí kořenu jen o málo menší než jedna, pak je vhodné zvolit jinou úpravu nebo použít jinou numerickou metodu.

Neexistuje žádný univerzální postup, jak lze rovnici $f(x) = 0$ upravit na tvar $x = g(x)$ tak, abychom dostali posloupnost konvergující ke kořenu. Má-li funkce f spojitou derivaci, často pomůže úprava na tvar $x = x - f(x)/k$, kde k je vhodná konstanta. Položíme-li $g(x) = x - f(x)/k$, platí $g'(x) = 1 - f'(x)/k$. Zvolíme-li za k hodnotu $k = f'(x_0)$, kde x_0 je číslo blízké hledanému kořenu, bude v jeho okolí platit $g'(x) \approx 0$ a je velká naděje, že metoda bude rychle konvergovat.

Podle znaménka a velikosti derivace $g'(x)$ mohou v okolí kořenu nastat v podstatě čtyři možnosti.

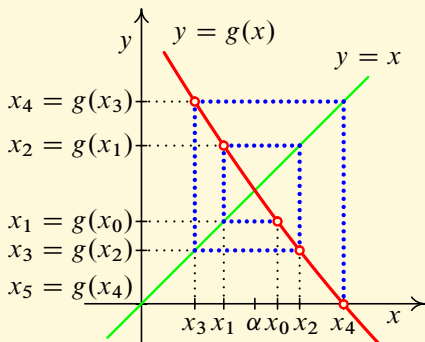
- (a) Posloupnost $\{x_n\}$ je monotónní a vzdaluje se (doleva nebo doprava) od kořenu α . Tato situace je znázorněna na obr. 2.12. Toto chování lze očekávat, když $1 < g'(\alpha)$.

Obr. 2.12: Prostá iterace — $\{x_n\}$ se monotónně vzdaluje doprava od kořenu α

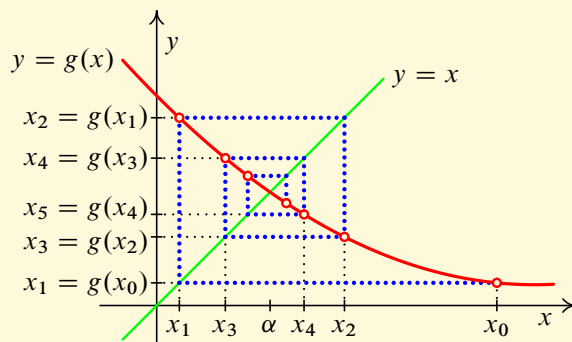


Obr. 2.13: Prostá iterace — $\{x_n\}$ monotónně konverguje zprava ke kořenu α

- (b) Posloupnost $\{x_n\}$ je monotónní a konverguje (zleva nebo zprava) ke kořenu α . Tato situace je znázorněna na obr. 2.13. Toto chování lze očekávat, když $0 < g'(\alpha) < 1$.
- (c) Posloupnost $\{x_n\}$ střídavě přeskakuje přes kořen α a vzdaluje se od něho. Tato situace je znázorněna na obr. 2.14. Toto chování lze očekávat, když $g'(\alpha) < -1$.
- (d) Posloupnost $\{x_n\}$ střídavě přeskakuje přes kořen α a konverguje k němu. Tato situace je znázorněna na obr. 2.15. Toto chování lze očekávat, když $-1 < g'(\alpha) < 0$.



Obr. 2.14: Prostá iterace — $\{x_n\}$ se vzdaluje střídavě vlevo a vpravo od kořenu α



Obr. 2.15: Prostá iterace — $\{x_n\}$ konverguje střídavě zleva a zprava ke kořenu α

Předchozí možnosti popisují chování posloupnosti $\{x_n\}$ v okolí kořenu α . Pokud tato posloupnost nekonverguje ke kořenu α , může být její chování velmi složité. V jednodušších případech je např. možné, že diverguje do nekonečna nebo konverguje k jinému kořenu. Může však také různě oscilovat apod.

Podrobnější informace o jednokrokových iteračních metodách a podmínkách jejich konvergence lze nalézt např. v [8, str. 188], [22, str. 26] nebo [52, str. 290].

Příklad 2.4 Metodou prosté iterace najděte nejmenší kladný kořen rovnice

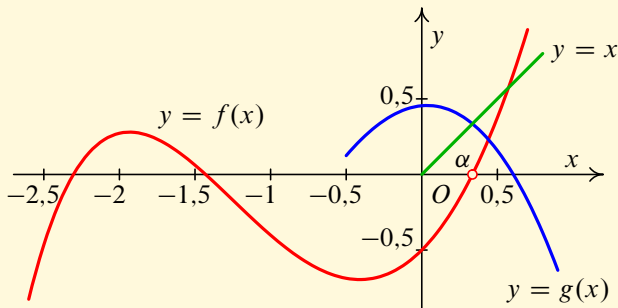
$$x^2 \ln(x + 4) + x - 0,5 = 0.$$

Použijte zastavovací podmínku $\frac{|x_n - x_{n-1}|}{|x_n|} < \varepsilon$, kde $\varepsilon = 0,05$.

Řešení. Nejprve nakreslíme graf funkce $f(x) = x^2 \ln(x + 4) + x - 0,5$ — viz obr. 2.16. Z obrázku je vidět, že hodnota nejmenšího kladného kořenu α je $\alpha \approx 0,3$.

Nyní musíme rovnici $f(x) = 0$ nějak upravit na tvar vhodný k iteraci. Použijeme lokálně ekvivalentní rovnici $x = x - \frac{f(x)}{1,1}$, za funkci $g(x)$, jejíž pevný bod budeme hledat, tudíž zvolíme $g(x) = x - \frac{f(x)}{1,1}$. Grafy této funkce a rovněž přímky $y = x$ jsou také znázorněné na obr. 2.16. Vlastně hledáme x -ovou souřadnici průsečíku těchto dvou grafů, což je právě kořen α .

Za počáteční aproximaci vybereme $x_0 = 0,06$. Pomocí vzorce (2.7) vypočítáme další členy x_1, x_2 atd. Výsledky jsou zapsány v následující tabulce.



Obr. 2.16: Graf funkce z příkladu 2.4

Je vidět, že zastavující podmínka byla splněna po šestnácti krocích. Platí $x_{16} = 0,328\,408\,829\,0$, tedy $\alpha \approx 0,328\,408\,829\,0$.

Na obr. 2.17 je uvedena animace, která znázorňuje konstrukci jednotlivých členů posloupnosti $\{x_n\}$, jež konverguje ke kořenu α . ▲

n	x_n	chyba
0	0,0600000000	—
1	0,4554143103	0,8682518343
2	0,2142342305	1,125777516
3	0,4140027827	0,4825294915
4	0,2608283384	0,5872615117
5	0,3886125134	0,3288215654
6	0,2868192317	0,3549039620
7	0,3717646516	0,2284924603
8	0,3029956456	0,2269636775
9	0,3602959011	0,1590366566
10	0,3135223314	0,1491873625
11	0,3524247390	0,1103850079
12	0,3205206557	0,09953830660
13	0,3470130449	0,07634407291
14	0,3252257322	0,6699135567
15	0,3432941626	0,05263250113
16	0,3284088290	0,04532561943

Obr. 2.17: Animace
k příkladu 2.4 —
metoda prosté iterace

2.2.5 Newtonova metoda (metoda tečen)

Tato metoda vyžaduje, aby existovala derivace $f'(x)$. Newtonova metoda totiž využívá k nalezení kořenu rovnice $f(x) = 0$ tečny ke grafu funkce f .

Postup:

Graf funkce f v okolí kořenu nahrazujeme tečnou. Předpokládejme, že existuje $f'(x)$ a že v nějakém okolí kořenu α platí $f'(x) \neq 0$.

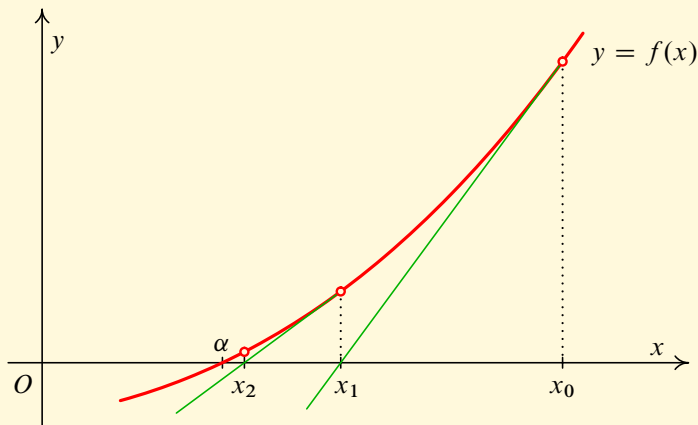
- (1) Zvolíme počáteční aproximaci x_0 , která leží dostatečně blízko kořenu α . V bodě $[x_0, f(x_0)]$ nyní sestrojíme tečnu ke grafu funkce $f(x)$. Průsečík tečny s osou x bude nová aproximace kořenu.

Rovnice tečny ke grafu funkce $f(x)$ v bodě $[x_0, f(x_0)]$ je

$$y = f(x_0) + f'(x_0)(x - x_0).$$

Najdeme průsečík x_1 této tečny s osou x , viz obr. 2.18. Z rovnice

$$0 = f(x_0) + f'(x_0)(x_1 - x_0)$$



Obr. 2.18: Princip Newtonovy metody

dostaneme, že

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

- (2) Analogicky postupujeme v dalších krocích. Obecně člen x_{n+1} sestrojíme z členu x_n tak, že sestrojíme tečnu ke grafu funkce $f(x)$ v bodě $[x_n, f(x_n)]$ a určíme její průsečík s osou x . Rovnice této tečny je

$$y = f(x_n) + f'(x_n)(x - x_n).$$

Z rovnice

$$0 = f(x_n) + f'(x_n)(x_{n+1} - x_n)$$

dostaneme, že

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots \quad (2.8)$$

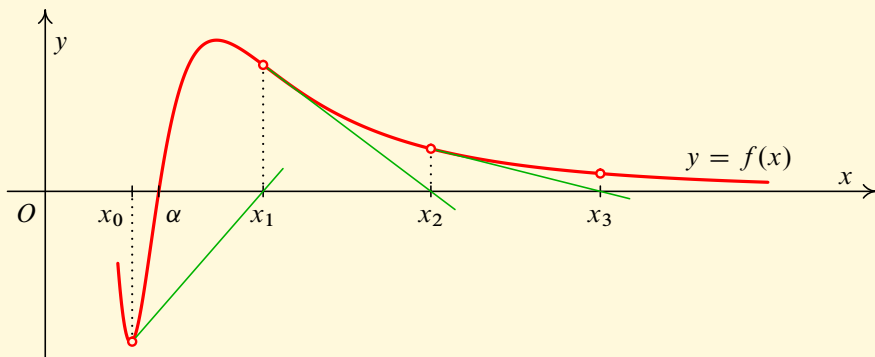
Nový člen x_{n+1} posloupnosti $\{x_n\}$ se tudíž počítá z předchozího členu x_n . Newtonova metoda je proto *stacionární jednokroková iterační metoda*.

Konvergence:

Metoda nemusí konvergovat. Lze však ukázat, že pokud má f spojitou druhou derivaci, α je jednoduchý kořen, tj. platí $f'(\alpha) \neq 0$, a počáteční aproximaci x_0 zvolíme dostatečně blízko kořenu, dává vztah (2.8) posloupnost $\{x_n\}$ konvergující ke kořenu α . Rychlost konvergence je přitom aspoň kvadratická. Viz [22, str. 41] nebo [52, str. 295]. Pro svou rychlost se tato metoda velmi často užívá. Je-li kořen vícenásobný, tj. $f'(\alpha) = 0$, je konvergence pouze lineární, viz [52, str. 295].

Konvergence je zaručena, platí-li tzv. *Fourierovy podmínky* $f(a) \cdot f(b) < 0$ (existuje kořen), $f'(x) \neq 0$ na intervalu $\langle a, b \rangle$ (funkce je ryze monotónní) a $f''(x) \neq 0$ na intervalu $\langle a, b \rangle$ (funkce je buď konvexní nebo konkávní). Za x_0 je třeba zvolit ten konec $c \in \{a, b\}$, pro nějž platí $f(c) \cdot f''(c) > 0$. Posloupnost $\{x_n\}$ je přitom monotónní. Viz [22, str. 44]. Na obr. 2.18 je funkce rostoucí a konvexní, takže platí $f(b) \cdot f''(b) > 0$. Správná volba je v tomto případě proto $x_0 = b$. Začneme-li v druhém konci, může se stát, že průsečík tečny s osou x je daleko od kořenu α .

Pokud nejsou splněny Fourierovy podmínky a počáteční aproximace x_0 není dostatečně blízko kořenu α , nemusí metoda konvergovat nebo může konvergovat k jinému kořenu, jak ilustruje obr. 2.19.



Obr. 2.19: Newtonova metoda může divergovat, nejsou-li splněny Fourierovy podmínky

Příklad 2.5 Newtonovou metodou najděte největší záporný kořen rovnice

$$1,9e^{-0,01x} - 0,1x - \sqrt{4 - x^2} = 0.$$

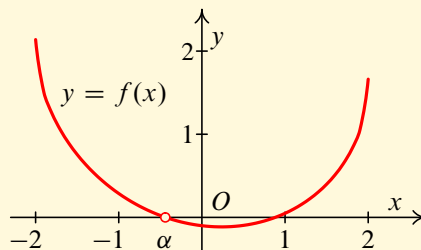
Použijte zastavovací podmínku $|x_n - x_{n-1}| < \varepsilon$, kde $\varepsilon = 0,001$.

Řešení. Nejprve nakreslíme graf funkce $f(x) = 1,9e^{-0,01x} - 0,1x - \sqrt{4 - x^2}$ — viz obr. 2.20. Z obrázku je vidět, že hodnota největšího záporného kořenu α je $\alpha \approx -0,5$.

Za počáteční aproximaci vybereme $x_0 = -1,95$. Pomocí vzorce (2.8) vypočítáme další členy x_1, x_2 atd. Výsledky jsou zapsány v sousední tabulce.

Je vidět, že zastavující podmínka byla splněna po šesti krocích. Platí $x_6 = -0,435\,986\,735\,8$, tedy $\alpha \approx -0,435\,986\,735\,8$.

Na obr. 2.21 je uvedena animace, která znázorňuje konstrukci jednotlivých členů posloupnosti $\{x_n\}$, jež konverguje ke kořenu α . ▲



Obr. 2.20: Graf funkce z příkladu 2.5

n	x_n	chyba
0	-1,950000000	—
1	-1,575488656	0,374511344
2	-0,9634337644	0,6120548916
3	-0,5716120111	0,3918217533
4	-0,4482688909	0,1233431202
5	-0,4361033189	0,0121655720
6	-0,4359867358	0,0001165831

Obr. 2.21: Animace
k příkladu 2.5 —
Newtonova metoda
(metoda tečen)

2.2.6 Metoda sečen

Newtonova metoda je velmi rychlá, ale vyžaduje v každém kroku výpočet derivace $f'(x_n)$. V případech, kdy je derivace dána příliš složitým vztahem nebo neexistuje, se proto často používá náhrada tečny sečnou. Sečnou rozumíme přímkou, která prochází dvěma body grafu funkce $f(x)$.

Postup:

Tato metoda vyžaduje dva startovací body x_0, x_1 , blízké kořenu α rovnice $f(x) = 0$. Nemusí nutně platit $f(x_0)f(x_1) < 0$, tedy kořen nemusí ležet mezi nimi. Myšlenka je obdobná jako u Newtonovy metody, jen místo tečny nahrazujeme graf funkce f sečnou.

- (1) Graf funkce f nahradíme přímkou, která prochází body $[x_0, f(x_0)]$ a $[x_1, f(x_1)]$. Lineární funkce l , jejímž grafem je tato přímka, má vzorec $l(x) = f(x_1) + k(x - x_1)$, kde $k = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$ je směrnice. Předpokládejme, že tato přímka protne osu x v bodě x_2 . Z rovnice

$$f(x_1) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x_2 - x_1) = 0$$

vypočítáme, že

$$x_2 = x_1 - \frac{x_1 - x_0}{f(x_1) - f(x_0)} f(x_1).$$

Předpokládáme, že jmenovatel zlomku v předchozím výrazu je nenulový.

Vlastně jsme oproti Newtonově metodě nahradili směrnicí $f'(x_1)$ tečny ke grafu funkce f v bodě $[x_1, f(x_1)]$ směrnicí sečny, která je určena dvěma body $[x_0, f(x_0)]$ a $[x_1, f(x_1)]$. Pro blízké body lze totiž předpokládat, že

$$f'(x_1) \approx \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

- (2) Postup opakujeme. Obecně v n -tém kroku nahradíme graf funkce f přímkou, která prochází body $[x_{n-1}, f(x_{n-1})]$ a $[x_n, f(x_n)]$. Lineární funkce l , jejímž grafem je tato přímka, má vzorec $l(x) = f(x_n) + k(x - x_n)$, kde $k = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$ je směrnice. Předpokládejme, že tato přímka protne osu x v bodě x_{n+1} . Z rovnice

$$f(x_n) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} (x_{n+1} - x_n) = 0$$

dostaneme, že

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n), \quad n = 0, 1, 2, \dots \quad (2.9)$$

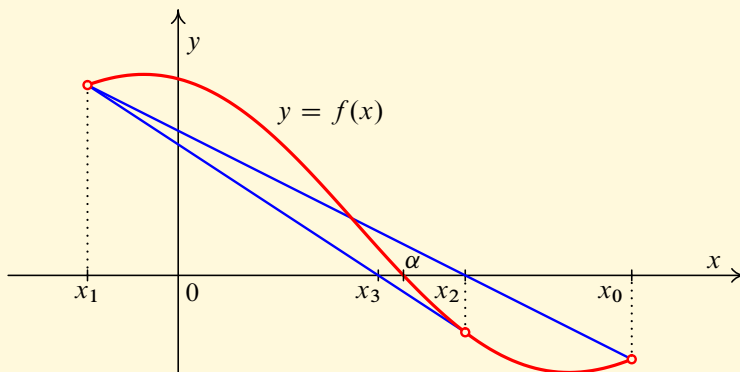
Předpokládáme, že jmenovatel zlomku ve vztahu (2.9) je nenulový pro libovolné $n = 0, 1, 2, \dots$

- (3) Nový člen x_{n+1} posloupnosti $\{x_n\}$ se tudíž počítá ze dvou předchozích členů x_{n-1} a x_n . Metoda sečen je proto *dvoukroková*. Říkáme, že jde o *stacionární dvoukrokovou iterační metodu*. Srovnejte se vzorcem (2.4) pro metodu regula falsi, která je nestacionární. Všimněte si rovněž, že na rozdíl od metody bisekce nebo regula falsi nemusí platit, že $f(x_{n-1})f(x_n) < 0$.

Na obr. 2.22 jsou znázorněny první dva kroky metody sečen.

Konvergence:

Metoda není vždy konvergentní. Konvergence je zaručena, pokud má funkce spojitou první a druhou derivaci, přičemž $f'(\alpha) \neq 0$ (kořen je jednoduchý), a pokud volíme počáteční aproximace x_0 a x_1 blízko kořenu α . Rychlost konvergence je za uvedených



Obr. 2.22: Princip metody sečen

předpokladů řádu $\frac{1}{2}(1 + \sqrt{5}) \doteq 1,618$ (toto číslo je kladným kořenem kvadratické rovnice $\mu^2 - \mu - 1 = 0$). Viz [22, str. 49] nebo [52, str. 341].

Metoda sečen má nižší řád než Newtonova metoda, takže lze očekávat, že konverguje pomaleji. Nicméně lze ukázat, že při srovnatelné výpočetní náročnosti (dva kroky metody

sečen jsou maximálně tak drahé jako jeden krok Newtonovy metody) lokálně konverguje rychlostí $\frac{1}{2}(1 + \sqrt{5}) + 1 \doteq 2,618$, tj. rychleji než Newtonova metoda, viz [52, str. 342].

Použití vyžaduje jistou opatrnost. Je-li zastavovací podmínka např. $|f(x_n)| < \varepsilon$, kde ε je velmi malé, může se stát, že podmínka ještě není splněna, avšak v rámci strojové přesnosti je $x_n = x_{n-1}$, takže ve vzorci (2.9) dostaneme neurčitý výraz $\frac{0}{0}$ a výpočet zhavaruje.

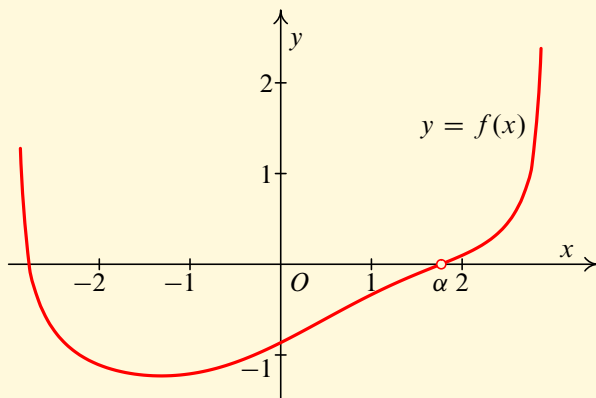
Příklad 2.6 Metodou sečen najděte nejmenší kladný kořen rovnice

$$\cos(0,3x)\sqrt{x^2 + 2x + 4} - 2(9 - 0,016x^6)^{0,25} + 0,6 = 0.$$

Použijte zastavovací podmínku $|x_n - x_{n-1}| < \varepsilon$, kde $\varepsilon = 0,05$.

Řešení. Nejprve nakreslíme graf funkce $f(x) = \cos(0,3x)\sqrt{x^2 + 2x + 4} - 2(9 - 0,016x^6)^{0,25} + 0,6$ — viz obr. 2.23. Z obrázku je vidět, že hodnota nejmenšího kladného kořenu α je $\alpha \approx 1,7$. Zvolíme $x_0 = 2,86$ a $x_1 = 2,8$.

Pomocí vzorce (2.9) vypočítáme další členy x_2, x_3 atd. Výsledky jsou zapsány v následující tabulce. Je vidět, že zastavující podmínka byla splněna po šesti krocích. Platí $x_7 = 1,768\,854\,775$, tedy $\alpha \approx 1,768\,854\,775$.



Obr. 2.23: Graf funkce z příkladu 2.6

n	x_n	chyba
0	2,860000000	—
1	2,800000000	—
2	2,693153896	0,106846104
3	2,499248330	0,193905566
4	2,254566608	0,244681722
5	1,992118060	0,262448548
6	1,814275893	0,177842167
7	1,768854775	0,045421118

Na obr. 2.24 je uvedena animace, která znázorňuje konstrukci jednotlivých členů posloupnosti $\{x_n\}$, jež konverguje ke kořenu α . Všimněte si, že na rozdíl od metody regula falsi je podstatné, který bod je označen jako x_0 a který jako x_1 . Při změně pořadí dostaneme jinou posloupnost — viz animace na obr. 2.25. ▲

Obr. 2.24: Animace
k příkladu 2.6 —
metoda sečen

Obr. 2.25: Animace

k příkladu 2.6 —
metoda sečen —
změna pořadí
startovacích bodů

Poznámka 2.7

- i) Kromě pěti klasických metod, které jsme uvedli, existuje mnoho dalších metod. Jejich popis lze najít v řadě učebnic numerické matematiky, např. [4, 22, 42, 43, 52].
- ii) Důležitou úlohou je hledání kořenů polynomů. Pro hledání reálných kořenů lze použít kteroukoli z předchozích metod. Vzhledem k významu této úlohy však byly vyvinuty speciální velmi účinné metody pro nalezení jak reálných tak i komplexních kořenů, viz např. [22, 52].

2.3 Systémy nelineárních rovnic

Obdobně jako se řeší numericky jedna rovnice o jedné neznámé, je možné řešit i soustavy rovnic. Předpokládejme, že je dán systém n rovnic o n neznámých

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0. \end{aligned}$$

Jeho kořenem rozumíme uspořádanou n -tici čísel $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_n]$ takovou, že platí rovnosti $f_i(\alpha_1, \dots, \alpha_n) = 0$ pro $i = 1, \dots, n$.

Označíme-li $\mathbf{x} = [x_1, \dots, x_n]^T$ a $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_n(\mathbf{x})]^T$, můžeme předchozí systém stručně zapsat pomocí vektorové symboliky jako $\mathbf{f}(\mathbf{x}) = \mathbf{0}$.

Obecně jde o náročnou problematiku, proto se omezíme pouze na základní informace. Výchozí myšlenka je stejná — nejprve se určí přibližná hodnota kořenu (separace) a pak se najde posloupnost $\{\mathbf{x}^{(k)}\}$ taková, že $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \boldsymbol{\alpha}$ (aproximace). Limitu chápeme po složkách, tedy označíme-li $\mathbf{x}^{(k)} = [x_1^{(k)}, \dots, x_n^{(k)}]^T$, musí platit $\lim_{k \rightarrow \infty} x_i^{(k)} = \alpha_i$ pro

$i = 1, \dots, n$. Podle důsledku 1.9 je to ekvivalentní konvergenci posloupnosti $\{\mathbf{x}^{(k)}\}$ k vektoru $\boldsymbol{\alpha}$ v libovolné vektorové normě.

Rovněž zastavovací podmínky se používají obdobně, jen místo absolutní hodnoty se použije vhodná vektorová norma. Tedy výpočet ukončujeme, když je splněna některá z podmínek

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| < \varepsilon, \quad \frac{\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|}{\|\mathbf{x}^{(k)}\|} < \varepsilon, \quad \|\mathbf{f}(\mathbf{x}^{(k)})\| < \varepsilon,$$

kde $\varepsilon > 0$ je zadané malé číslo.

Separace kořenů

Nalezení přibližné polohy kořenů je velmi obtížné. Neexistuje žádný univerzální postup. Pro $n = 2$ lze odhad provést graficky pomocí vhodného softwaru (znázorníme „křivky“, které odpovídají rovnicím $f_1(x_1, x_2) = 0$, $f_2(x_1, x_2) = 0$ a odhadneme souřadnice jejich průsečíků). Někdy lze kořen odhadnout na základě znalosti konkrétního problému, jehož matematickým modelem soustava je.

Aproximace kořenů

Ne všechny metody pro nalezení kořenu rovnice o jedné neznámé lze přenést na systémy. Nejde to např. u těch, které jsou založené na tom, že spojitá funkce má mezi dvěma body, v nichž nabývá hodnot opačných znamének, aspoň jeden kořen. Z těch, které jsme uvedli, je to metoda bisekce a regula falsi. Zmíníme se pouze o metodě prosté iterace a Newtonově metodě pro soustavy rovnic. Kromě nich existuje řada dalších, důležité a účinné jsou zejména metody, které úlohu najít kořen převedou na hledání minima funkce více proměnných. Vzhledem k rozsahu a určení tohoto textu se jimi nemůžeme zabývat. Viz např. [8, 22, 43, 52].

Iterační metody mají ve vektorovém zápisu obdobný tvar jako v případě jedné rovnice, a to

$$\mathbf{x}^{(k+1)} = \mathbf{g}(\mathbf{x}^{(k)}), \quad k = 0, 1, \dots,$$

kde $\mathbf{x}^{(k)} = [x_1^{(k)}, \dots, x_n^{(k)}]^T$ je k -tá aproximace kořenu $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$ a $\mathbf{g} = [g_1, g_2, \dots, g_n]$ je n -tice funkcí n proměnných.

V dalším výkladu se pro jednoduchost omezíme na případ dvou rovnic o dvou neznámých, které označíme x a y . Obecný případ je obdobný. Protože nebudeme potřebovat

dolní indexy k označení neznámých, použijeme je k označení pořadí iterace, čímž se značení zjednoduší. Tudíž $[x_k, y_k]$ bude značit k -tou aproximaci kořenu, který označíme $[\alpha, \beta]$.

2.3.1 Metoda prosté iterace

Budeme řešit soustavu dvou rovnic

$$f_1(x, y) = 0,$$

$$f_2(x, y) = 0$$

mající kořen $[\alpha, \beta]$, tj. platí, že $f_1(\alpha, \beta) = 0$ a $f_2(\alpha, \beta) = 0$. Musíme ji nahradit ekvivalentní soustavou (aspoň v okolí hledaného kořenu)

$$x = g_1(x, y),$$

$$y = g_2(x, y),$$

pro kterou platí, že $\alpha = g_1(\alpha, \beta)$ a $\beta = g_2(\alpha, \beta)$.

Zvolíme počáteční aproximaci $[x_0, y_0]$. Pak položíme

$$x_1 = g_1(x_0, y_0),$$

$$y_1 = g_2(x_0, y_0).$$

Obdobně pokračujeme dál. Obecně bude platit

$$\begin{aligned} x_{k+1} &= g_1(x_k, y_k), \\ y_{k+1} &= g_2(x_k, y_k), \end{aligned} \quad k = 0, 1, 2, \dots \quad (2.10)$$

Podmínky zaručující konvergenci $\lim_{k \rightarrow \infty} x_k = \alpha$ a $\lim_{k \rightarrow \infty} y_k = \beta$ jsou poměrně komplikované, viz např. [22, str. 64] nebo [52, str. 293]. Zhruba lze říci, že pokud mají funkce g_1 a g_2 spojitě první parciální derivace, jejichž hodnoty jsou v okolí kořenu malé, a počáteční aproximace $[x_0, y_0]$ je dostatečně blízko kořenu, bude iterační proces (2.10) konvergovat k řešení $[\alpha, \beta]$.

2.3.2 Newtonova metoda pro systémy

Budeme řešit soustavu dvou rovnic

$$\begin{aligned}f_1(x, y) &= 0, \\f_2(x, y) &= 0.\end{aligned}\tag{2.11}$$

Rovnice linearizujeme v okolí bodu $[x_0, y_0]$, který leží dostatečně blízko přesného řešení, které hledáme.

Připomeňme definici totálního diferenciálu funkce g , mající spojité první parciální derivace. Označíme-li $h = x - x_0$ a $k = y - y_0$, platí

$$g(x_0 + h, y_0 + k) = g(x_0, y_0) + \frac{\partial g}{\partial x}(x_0, y_0) \cdot h + \frac{\partial g}{\partial y}(x_0, y_0) \cdot k + R(h, k)$$

resp. při jiném označení

$$\begin{aligned}g(x, y) &= g(x_0, y_0) + \frac{\partial g}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial g}{\partial y}(x_0, y_0) \cdot (y - y_0) + \\&+ R(x - x_0, y - y_0),\end{aligned}$$

kde zbytek R je (v rozumných případech) malý pro $[h, k] = [x - x_0, y - y_0]$ blízké $[0, 0]$. Jeho zanedbáním dostaneme, že pro $[x, y]$ blízké $[x_0, y_0]$ přibližně platí

$$g(x, y) \approx g(x_0, y_0) + \frac{\partial g}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial g}{\partial y}(x_0, y_0) \cdot (y - y_0).$$

Výraz napravo je lineární funkce proměnných x, y , která je rovnicí tečné roviny ke grafu funkce g v bodě $[x_0, y_0, g(x_0, y_0)]$.

Tímto způsobem nahradíme obě funkce v soustavě (2.11). Dostaneme následující přibližné vztahy:

$$f_1(x, y) \approx f_1(x_0, y_0) + \frac{\partial f_1}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial f_1}{\partial y}(x_0, y_0) \cdot (y - y_0),$$

$$f_2(x, y) \approx f_2(x_0, y_0) + \frac{\partial f_2}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial f_2}{\partial y}(x_0, y_0) \cdot (y - y_0).$$

Místo soustavy (2.11) budeme řešit pomocnou lineární soustavu (neznámé jsou x a y)

$$\begin{aligned} f_1(x_0, y_0) + \frac{\partial f_1}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial f_1}{\partial y}(x_0, y_0) \cdot (y - y_0) &= 0, \\ f_2(x_0, y_0) + \frac{\partial f_2}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial f_2}{\partial y}(x_0, y_0) \cdot (y - y_0) &= 0. \end{aligned} \quad (2.12)$$

Přepíšeme ji do maticového zápisu. Zavedeme označení

$$\mathbf{f}(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix}, \quad \mathbf{J}(x, y) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(x, y) & \frac{\partial f_1}{\partial y}(x, y) \\ \frac{\partial f_2}{\partial x}(x, y) & \frac{\partial f_2}{\partial y}(x, y) \end{pmatrix}.$$

Tedy \mathbf{f} je sloupcová matice 2×1 a \mathbf{J} (tzv. *Jacobiova¹ matice*) je čtvercová 2×2 . Pak soustavu (2.11) lze stručně zapsat jako

$$\mathbf{f}(x, y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

¹**Carl Gustav Jacob Jacobi** (1804–1851) (čti jakobi) — významný německý matematik. Zabýval se teorií funkcí, lineární algebrou, diferenciálními rovnicemi a mechanikou.

a její linearizovanou verzi (2.12) (s využitím maticových operací sečítání a násobení) jako

$$\mathbf{f}(x_0, y_0) + \mathbf{J}(x_0, y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (2.13)$$

Pokud existuje inverzní matice $\mathbf{J}^{-1}(x_0, y_0)$, soustavu (2.13) snadno vyřešíme:

$$\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = -\mathbf{J}^{-1}(x_0, y_0) \mathbf{f}(x_0, y_0),$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \mathbf{J}^{-1}(x_0, y_0) \mathbf{f}(x_0, y_0).$$

Označme její řešení x_1 a y_1 , tj.

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \mathbf{J}^{-1}(x_0, y_0) \mathbf{f}(x_0, y_0).$$

Nyní nahradíme bod $[x_0, y_0]$ bodem $[x_1, y_1]$ a obdobně určíme bod $[x_2, y_2]$. A tak dál. Obecně bude platit:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \mathbf{J}^{-1}(x_k, y_k) \mathbf{f}(x_k, y_k), \quad k = 0, 1, 2, \dots \quad (2.14)$$

Poznámka 2.8 Porovnejte vzorec (2.14) se vzorcem (2.8) pro „skalární“ případ

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)},$$

který jsme použili pro řešení jedné rovnice $f(x) = 0$ Newtonovou metodou (náhrada grafu funkce f tečnou v bodě $[x_k, f(x_k)]$). Zřejmě jde o jeho maticovou verzi. V tomto případě je totiž $\mathbf{J}(x) = (f'(x))$, takže tato jednorozměrná matice je regulární, právě když $f'(x) \neq 0$. Inverzní matice je pak $\mathbf{J}^{-1}(x) = (1/f'(x))$.

Obdobně lze postupovat pro soustavy tří a více rovnic, maticový zápis vzorce pro výpočet iterací je obdobný jako (2.14).

Výpočet iterace $[x_{k+1}, y_{k+1}]$ z iterace $[x_k, y_k]$ pomocí vzorce (2.14) vyžaduje výpočet inverzní matice $\mathbf{J}^{-1}(x_k, y_k)$, což není problém pro dvě neznámé, ale pro větší počet je to numericky poměrně náročné. Proto se výpočet obvykle organizuje jinak. Označme

$$\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} - \begin{pmatrix} x_k \\ y_k \end{pmatrix}.$$

Pak ze (2.14) je vidět, že

$$\mathbf{J}(x_k, y_k) \begin{pmatrix} u_k \\ v_k \end{pmatrix} = -\mathbf{f}(x_k, y_k),$$

což je soustava lineárních rovnic s neznámými u_k a v_k . Na nalezení jejího řešení existují velmi účinné numerické metody, viz kapitola 3. Potom

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \begin{pmatrix} u_k \\ v_k \end{pmatrix}.$$

Tento postup je (zejména pro větší počet neznámých) daleko účinnější.

Konvergence:

Lze ukázat, že pokud funkce f_1 a f_2 mají spojité druhé parciální derivace, platí, že $\det \mathbf{J}(x, y) \neq 0$ v blízkosti přesného řešení soustavy (2.11) (tato podmínka zajišťuje existenci inverzní matice k Jacobiově matici) a počáteční aproximace $[x_0, y_0]$ je dostatečně blízká tomuto řešení, pak posloupnost $[x_k, y_k]$ daná vzorcem (2.14) konverguje pro $k \rightarrow \infty$ k přesnému řešení a rychlost konvergence je kvadratická. Obdobný výsledek

platí i pro soustavy o více neznámých. Přesná formulace je poměrně komplikovaná, viz např. [4, II. díl, str. 155], [30, str. 703], [52, str. 299] nebo [69, str. 20].

Poznámka 2.9 Newtonovu metodu lze dalekosáhle zobecnit na operátorové rovnice, které v sobě zahrnují jako speciální případ nejen výše uvažované rovnice v \mathbb{R}^n , ale rovněž obyčejné a parciální diferenciální rovnice apod. Viz [8, str. 290] nebo [30, str. 689]. Touto problematikou se zabýval Kantorovič¹.

Pojmy k zapamatování

- kořen nelineární rovnice
- existence a jednoznačnost kořenu
- separace kořenů
- aproximace kořenů

¹**Leonid Vitaljevič Kantorovič** (1912–1986) — ruský matematik. Zabýval se funkcionální analýzou a matematickými metodami v ekonomii (matematické programování). V roce 1975 obdržel Nobelovu cenu za ekonomii (společně s **Tjallingem C. Koopmansem**) za „příspěvek k teorii optimální alokace zdrojů“.

- zastavovací podmínky
- rychlost konvergence, řád iterační metody
- startovací a zpřesňující metody
- metoda bisekce (půlení intervalu)
- metoda regula falsi (tětiv)
- metoda prosté iterace
- pevný bod funkce
- Newtonova metoda (tečen)
- metoda sečen
- kořen systému nelineárních rovnic
- separace a aproximace kořenů systémů rovnic
- metoda prosté iterace pro systémy
- Newtonova metoda pro systémy

Kontrolní otázky

1. Která čísla nazýváme kořeny rovnice $f(x) = 0$?
2. Jakými způsoby lze získat informaci o poloze kořenů rovnice $f(x) = 0$?
3. Popište postup při hledání kořenů.
4. Vysvětlete, co znamená pojem separace kořenů.
5. Vysvětlete, co znamená pojem aproximace kořenů.
6. Uveďte podmínky, které zaručují, že funkce $f(x) = 0$ má na intervalu $\langle a, b \rangle$ alespoň jeden kořen.
7. Pro funkci $f(x)$ spojitou na intervalu $\langle a, b \rangle$ platí $f(a)f(b) < 0$. Co můžeme říci o počtu kořenů rovnice $f(x) = 0$?
8. Pro funkci $f(x)$ spojitou na intervalu $\langle a, b \rangle$ platí $f(a)f(b) < 0$, přičemž $f'(x)$ nemění na (a, b) znaménko. Co můžeme říci o počtu kořenů rovnice $f(x) = 0$?
9. Uveďte podmínku, která zaručuje, že spojitá funkce splňující $f(a)f(b) < 0$ má na intervalu $\langle a, b \rangle$ právě jeden kořen.
10. Uveďte podmínky, které se používají pro zastavení výpočtu při hledání kořenu numerickou metodou.

11. V případě, že je splněna některá ze zastavovacích podmínek na straně 120, znamená to, že platí $|\alpha - x_n| < \varepsilon$?
12. Jaké jsou přednosti a nedostatky jednotlivých zastavovacích podmínek na straně 120?
13. Vysvětlete, co je řád konvergence posloupnosti $\{x_n\}$ a řád iterační metody.
14. Vysvětlete princip metody bisekce.
15. Kolik startovacích bodů potřebuje metoda bisekce?
16. Vysvětlete princip metody regula falsi.
17. Co rozumíme tětivou u metody regula falsi?
18. Kolik startovacích bodů potřebuje metoda regula falsi?
19. Uveďte, jaké podmínky zaručují, že metoda regula falsi konverguje.
20. Vysvětlete princip metody prosté iterace.
21. S čím souvisí rychlost konvergence prosté iterační metody?
22. Jaké podmínky zaručují, že metoda prosté iterace konverguje?
23. Vysvětlete princip Newtonovy metody.
24. Kolik startovacích bodů potřebuje Newtonova metoda?
25. Uveďte, jaké podmínky zaručují, že Newtonova metoda konverguje.

26. Kdy je výhodnější místo Newtonovy metody použít metodu sečen?
27. Vysvětlete princip metody sečen.
28. Co rozumíme sečnou u metody sečen?
29. Kolik startovacích bodů potřebuje metoda sečen?
30. Uveďte, jaké podmínky zaručují, že metoda sečen konverguje.
31. Která z probíraných metod používajících dva startovací body nevyžaduje, aby funkční hodnoty ve startovacích bodech měly opačná znaménka?
32. Které z iteračních metod pro nalezení kořenů jsou vždy konvergentní, pokud spojitá funkce $f(x)$ na intervalu $\langle a, b \rangle$ splňuje podmínku $f(a)f(b) < 0$?
33. Které metody označujeme jako startovací a proč?
34. Které metody označujeme jako zpřesňující a proč?
35. Co rozumíme kořenem systému rovnic?
36. Jaké znáte iterační metody pro systémy nelineárních rovnic?
37. Popište, jak lze postupovat při hledání kořenu systému dvou nelineárních rovnic.
38. Vysvětlete princip metody prosté iterace pro systémy rovnic.

39. Máte nalézt řešení soustavy rovnic $f_1(x, y) = 0$, $f_2(x, y) = 0$ pomocí metody prosté iterace. Napište, jak vypadá její přepis na ekvivalentní soustavu vhodnou pro řešení touto metodou.
40. Jaké podmínky zaručují, že metoda prosté iterace pro systémy konverguje?
41. Vysvětlete princip Newtonovy metody pro systémy rovnic.
42. Jaké podmínky zaručují, že Newtonova metoda pro systémy konverguje?

Testy ke kapitole 2

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Nechť funkce $f(x)$ má na intervalu $\langle a, b \rangle$ kořen. Aby bylo zaručeno, že tento kořen je jediný, stačí, aby
 - funkce $f(x)$ byla na tomto intervalu konkávní.
 - funkce $f(x)$ byla na tomto intervalu konvexní.
 - funkce $f(x)$ měla první derivaci rovnou nule.
 - funkce $f(x)$ byla na tomto intervalu ryze monotónní.

2. (1b.) Má kvadratická rovnice $(x - 1)^2 + 2 = 0$ reálné kořeny?

Nemá.

Ano, má kořeny $x_1 = 1$ a $x_2 = -2$.

Ano, má dvojnásobný kořen
 $x_{1,2} = 1 + \sqrt{-2}$.

Ano, má kořeny $x_{1,2} = 1 \pm \sqrt{-2}$.

3. (1b.) Máme nalézt největší záporný kořen rovnice $f(x) = 0$ pomocí Newtonovy metody. Pak

volíme libovolný startovací bod kdekoli na ose x .

volíme jeden startovací bod poblíž tohoto kořenu.

volíme dva startovací body tak, aby hledaný kořen ležel mezi nimi.

volíme dva startovací body tak, aby hledaný kořen ležel mezi nimi a funkční hodnoty v těchto bodech měly opačná znaménka.

4. (1b.) Vyberte, o kterou iterační metodu se jedná: Graf funkce $f(x)$ se na intervalu $\langle a, b \rangle$ nahradí přímkou, která prochází body $[a, f(a)]$ a $[b, f(b)]$, průsečík x_0 této přímky s osou x je přibližný odhad kořenu. Dále pak vybereme ten z intervalů $\langle a, x_0 \rangle$ nebo $\langle x_0, b \rangle$, pro který platí $f(a)f(x_0) < 0$ nebo $f(x_0)f(b) < 0$, a postup opakujeme.

Metoda bisekce.

Newtonova metoda (metoda tečen).

Metoda regula falsi (metoda třetiv).

Metoda sečen.

5. (1b.) V případě, že posloupnost čísel x_0, x_1, \dots konverguje ke kořenu α rovnice $f(x) = 0$ velmi pomalu, pak

zastavovací podmínku zvolíme raději nastavením maximálního počtu iterací.

zastavovací podmínku volíme ve tvaru $|x_n - x_{n-1}| < \varepsilon$.

zastavovací podmínku volíme ve tvaru $\frac{|x_n - x_{n-1}|}{|x_n|} < \varepsilon$.

6. (1b.) Je číslo 0 kořenem rovnice $2 \sin x + \frac{3}{5} e^x - 0,6 = x^2$?

Ne.

Ano.

Nelze rozhodnout.

7. (1b.) Jestliže rovnici $f(x) = 0$ přepíšeme do tvaru $f_1(x) = f_2(x)$, pak

funkční hodnoty v kořenech původní rovnice jsou stejné jako funkční hodnoty v x -ových souřadnicích průsečíků grafů funkcí $f_1(x)$ a $f_2(x)$.

funkční hodnoty v kořenech původní rovnice jsou stejné jako y -ové souřadnice průsečíků grafů funkcí $f_1(x)$ a $f_2(x)$.

kořeny původní rovnice a x -ové souřadnice průsečíků grafů funkcí $f_1(x)$ a $f_2(x)$ mají stejné hodnoty.

kořeny původní rovnice mají jiné hodnoty než x -ové souřadnice průsečíků grafů funkcí $f_1(x)$ a $f_2(x)$.

8. (1b.) Jsou-li splněny podmínky pro konvergenci prosté iterační metody $x_{n+1} = g(x_n)$ na intervalu $\langle a, b \rangle$, pak počet iterací nutných k dosažení předepsané přesnosti je ovlivněn znaménkem funkční hodnoty v počáteční aproximaci.

je ovlivněn volbou počáteční aproximace.

není ovlivněn vlastnostmi funkce $g(x)$.

lze ovlivnit tím, že funkci derivujeme a hledáme kořen rovnice $g'(x) = 0$.

9. (1b.) Co je kořenem soustavy n rovnic o n neznámých tvaru $f_i(x_1, \dots, x_n) = 0$, $i = 1, \dots, n$, $n > 1$?

Kořenem je reálné číslo α takové, že pro všechna $i = 1, \dots, n$ platí rovnosti $f_i(\alpha, \dots, \alpha) = 0$.

Kořenem je uspořádaná n -tice čísel $\alpha = [\alpha_1, \dots, \alpha_n]$ taková, že pro všechna $i = 1, \dots, n$ platí rovnosti $f_i(\alpha_i, \dots, \alpha_i) = 0$.

Kořenem je uspořádaná n -tice čísel $\alpha = [\alpha_1, \dots, \alpha_n]$ taková, že alespoň pro jedno $i = 1, \dots, n$ platí rovnost $f_i(\alpha_1, \dots, \alpha_n) = 0$.

Kořenem je uspořádaná n -tice čísel $\alpha = [\alpha_1, \dots, \alpha_n]$ taková, že pro všechna $i = 1, \dots, n$ platí rovnosti $f_i(\alpha_1, \dots, \alpha_n) = 0$.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (1b.) Kořenem rovnice $f(x) = 0$ je libovolné reálné číslo α , pro které platí:

$$\alpha = 0. \quad f(\alpha) = 0. \quad f(0) = \alpha. \quad f(\alpha) = 1.$$

2. (1b.) Polynom $P(x) = ax^3 - bx + c$, kde $a, b, c \in \mathbb{R}$, $a \neq 0$, má v reálném oboru právě 3 kořeny (počítáno včetně násobností).

má v reálném oboru vždy 3 různé kořeny.

má v komplexním oboru více než 3 kořeny.

má v komplexním oboru právě 3 kořeny (počítáno včetně násobností).

3. (1b.) Jestliže separujeme kořeny, pak

stanovíme jejich přesnou hodnotu pomocí některé z iteračních metod.

určíme intervaly, které jsou dostatečně malé a obsahují jediný kořen.

nalezneme posloupnost čísel x_0, x_1, \dots , která konverguje ke kořenu.

z obrázku přibližně odhadneme jejich hodnotu.

4. (1b.) Vyberte vhodný interval, kterým lze odstartovat metodu bisekce při hledání záporného kořenu funkce $f(x) = 3x^2 - 3$.

$\langle -2, 0 \rangle$.

$\langle -2, 2 \rangle$.

$\langle 0, 2 \rangle$.

$\langle 1/2, 3/2 \rangle$.

5. (1b.) Která z metod využívá při hledání kořenu rovnice $f(x) = 0$ ekvivalentní zápis $x = g(x)$?

Metoda bisekce.

Newtonova metoda.

Metoda prosté iterace.

Metoda sečen.

6. (1b.) Předpokládejme, že data v tabulce jsou ze spojitě funkce. Vyberte dvojici bodů, mezi kterými musí ležet alespoň jeden kořen této funkce.

x_i	1,0	1,5	2,0	2,3	2,7	3,0
$f(x_i)$	0,7	0,1	-0,4	-1,0	-2,1	-2,3

1,0 a 1,5.

1,5 a 2,0.

2,7 a 3,0.

2,3 a 3,0.

7. (1b.) Hledáme-li kořeny funkce $f(x)$ na intervalu $\langle a, b \rangle$ a vynecháme požadavek na spojitost funkce $f(x)$, pak

spojitost první derivace funkce $f(x)$ zaručuje, že je na intervalu $\langle a, b \rangle$ kořen.

v případě, že funkce $f(x)$ je na intervalu $\langle a, b \rangle$ monotónní a platí $f(a)f(b) < 0$, bude mít funkce na intervalu $\langle a, b \rangle$ kořen.

podmínka $f(a)f(b) < 0$ nezaručuje, že na intervalu $\langle a, b \rangle$ má funkce kořen.

podmínka $f(a)f(b) < 0$ je postačující k tomu, aby funkce $f(x)$ měla na intervalu $\langle a, b \rangle$ kořen.

8. (1b.) Vyberte iterační metodu pro hledání kořenů nelineárních rovnic, která potřebuje k odstartování pouze jeden bod blízko kořenu.

Metoda třítiv.

Metoda Newtonova.

Metoda sečen.

Metoda bisekce.

9. (1b.) Které z metod pro nalezení kořenu nelineární rovnice o jedné neznámé lze přenést na systémy n nelineárních rovnic o n neznámých?

Metodu prosté iterace a Newtonovu metodu.

Metodu bisekce a metodu třítiv.

Metodu prosté iterace a metodu sečen.

Newtonovu metodu a metodu regula falsi.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (1b.) Jaké podmínky zaručují, aby funkce $f(x)$ měla uvnitř intervalu $\langle a, b \rangle$ kořen?

Funkce $f(x)$ má v krajních bodech intervalu hodnoty opačných znamének.

Funkce $f(x)$ je na tomto intervalu spojitá.

Funkce $f(x)$ je na tomto intervalu spojitá a hodnoty v krajních bodech intervalu mají opačná znaménka.

Funkce $f(x)$ je na tomto intervalu spojitá a hodnoty v krajních bodech intervalu mají stejná znaménka.

2. (1b.) Vyberte tu odpověď, která obsahuje všechny kořeny polynomu

$$P(x) = x^4 + x^3 - 2x^2.$$

0, -1, 2.

0 (dvojnásobný kořen), 1, -2.

0, 1 (dvojnásobný kořen), -2.

0, -1, 1, -2.

3. (1b.) Aproximací kořenu rozumíme

nalezení posloupnosti čísel x_0, x_1, \dots , která konverguje ke kořenu α , tj. platí

$$\lim_{n \rightarrow \infty} x_n = \alpha, \quad f(\alpha) = 0.$$

nalezení intervalů, v nichž vždy leží jediný kořen.

nalezení čísel α , pro která platí $f(\alpha) = 0$.

postup, kterým nalezneme kořeny dané rovnice.

4. (1b.) Iterační metodou byla sestrojena posloupnost konvergující ke kořenu rovnice $f(x) = 0$. Výpočet byl zastaven po splnění podmínky $|x_{n+1} - x_n| < 5 \cdot 10^{-5}$. Určete, která posloupnost byla získána.

0,5; 0,10981; 0,20970; 0,27240; 0,28611; 0,28607; 0,28607.

0,9; -0,68658; 0,33871; 0,82083; -0,12534; 0,97675; -0,63310; 0,43512.

0,6; 0,41788; 0,72451; 0,13585; 0,97188; -0,61558; 0,46515.

0; 1; 0,73212; 0,37303; 0,58097; 0,55938; 0,55653; 0,55649.

5. (1b.) Kdy je zaručeno, že Newtonova metoda bude na intervalu $\langle a, b \rangle$ konvergovat?

Pokud má funkce $f(x)$ spojitou druhou derivaci a počáteční aproximaci x_0 zvolíme blízko kořenu.

Pokud jsou splněny podmínky $f(a)f(b) < 0$, $f'(x) \neq 0$, $f''(x) \neq 0$ a za počáteční aproximaci x_0 zvolíme a nebo b tak, aby $f(x_0)f''(x_0) > 0$.

Pokud jsou splněny podmínky $f(a)f(b) > 0$ a $f'(x) \neq 0$.

Newtonova metoda je vždy konvergentní.

6. (1b.) Na jakém principu funguje metoda bisekce?

Najdeme střed x_0 intervalu $\langle a, b \rangle$, zjistíme, zda je v něm funkční hodnota rovna nule, pokud ano, máme kořen. Jinak pokračujeme stejným způsobem dále s tím, že pracujeme s intervalem, pro který platí $f(a)f(x_0) > 0$ resp. $f(b)f(x_0) > 0$.

Najdeme střed x_0 intervalu $\langle a, b \rangle$, zjistíme, zda je v něm funkční hodnota rovna nule, pokud ano, máme kořen. Jinak pokračujeme stejným způsobem dále s tím, že pracujeme s intervalem pro který platí $f(a)f(x_0) < 0$ resp. $f(b)f(x_0) < 0$.

Najdeme střed x_0 intervalu $\langle a, b \rangle$, zjistíme, zda je v něm funkční hodnota rovna nule, pokud ano, máme kořen. Jinak pokračujeme stejným způsobem dále s tím, že pracujeme s libovolným z intervalů $\langle f(a), f(x_0) \rangle$ resp. $\langle f(x_0), f(b) \rangle$.

Ani jedna z předchozích možností neodpovídá metodě bisekce.

7. (1b.) Máme nalézt kořeny rovnice $x^3 - 2x^2 - 47x - 44 = 0$. Místo toho budeme hledat x -ové souřadnice průsečíků grafů funkcí $f_1(x) = x^3 - 2x^2$ a $f_2(x) = -47x - 44$. Najdeme stejné kořeny?

Ne.

Ano.

8. (1b.) Která z iteračních metod pro hledání kořenů na intervalu $\langle a, b \rangle$ je založena na tzv. větě o pevném bodu?

Prostá iterační metoda.

Metoda tečen (Newtonova metoda).

Metoda sečen.

Metoda třetiv (metoda regula falsi).

9. (1b.) Která z metod pro hledání kořenů nelineární rovnice $f(x) = 0$ vyžaduje existenci derivace funkce $f(x)$?

Prostá iterační metoda.

Metoda třetiv.

Metoda sečen.

Newtonova metoda.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Kapitola 3

Numerické řešení systémů lineárních rovnic

Cíle

Po prostudování této kapitoly budete schopni vysvětlit:

- jak vypadá maticový zápis systému lineárních rovnic,
- v čem spočívá obtížnost numerického řešení systémů lineárních rovnic,
- podle čeho metody klasifikujeme,
- které jsou nejdůležitější přímé metody a jak fungují,
- které jsou nejdůležitější iterační metody a jak fungují,
- co ovlivňuje velikost chyb při numerickém řešení systémů lineárních rovnic.

Řešení systémů lineárních rovnic patří k nejdůležitějším úlohám numerické matematiky. Důvodem je, že řada jiných problémů se převádí na tuto úlohu. Zdálnivě jde o jednoduchou záležitost. Ze základního kurzu lineární algebry víme, že na rozdíl od nelineárních rovnic dokážeme po konečném počtu kroků teoreticky určit přesné řešení lineárního systému. Skutečnost je ovšem poněkud jiná. V současnosti se řeší systémy mající řádově až miliony rovnic a neznámých. Výpočty takového rozsahu je pochopitelně nutné provádět numericky na počítači. Řešení proto dostaneme jen přibližně. Nevhodně zvolený postup může kvůli zaokrouhlovacím chybám u konkrétního systému způsobit, že dostaneme naprosto chybné řešení. Potíže tohoto druhu, způsobené výpočty v množině strojových čísel, lze ale demonstrovat i na zcela malých systémech, viz příklad 1.1.

Zformulujme nejprve přesně, jakou úlohu vlastně budeme řešit. Z lineární algebry je známo, že systém lineárních rovnic má buď jediné řešení, nebo žádné řešení, nebo nekonečně mnoho řešení. My se budeme zabývat numerickým řešením systémů majících jediné řešení. V tomto případě musí být počet rovnic stejný jako počet neznámých nebo větší. Je-li větší, jsou rovnice lineárně závislé a některé z nich je možné vynechat, takže zbylé rovnice budou lineárně nezávislé a bude jich tolik, kolik je neznámých.

Budeme tedy uvažovat čtvercový systém lineárních rovnic o n neznámých tvaru

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ \dots & \\a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n,\end{aligned}\tag{3.1}$$

kde $n \geq 1$ je přirozené číslo. Reálná čísla a_{ij} , $i = 1, \dots, n$, $j = 1, \dots, n$, jsou koeficienty u neznámých x_j , $j = 1, \dots, n$, a reálná čísla b_i , $i = 1, \dots, n$, jsou koeficienty pravých stran. Označíme

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Pak $A \in \mathbb{M}_n$ je *matice soustavy*, $B \in \mathbb{R}^n$ je *vektor pravých stran* a $\mathbf{x} \in \mathbb{R}^n$ je *vektor neznámých*. Systém (3.1) pak můžeme stručně maticově zapsat ve tvaru

$$A\mathbf{x} = B,\tag{3.2}$$

kde $A\mathbf{x}$ je maticový součin čtvercové matice A a sloupcové matice \mathbf{x} . Předpoklad, že tento systém má právě jedno řešení, je ekvivalentní tomu, že matice A je *regulární*, tj. $\det A \neq 0$. Tato podmínka je rovněž ekvivalentní s existencí inverzní matice A^{-1} , tedy matice, pro niž platí $AA^{-1} = A^{-1}A = E$, kde E je jednotková matice. Další ekvivalentní podmínkou je, že pro hodnotu matice A platí $h(A) = n$. Připomeňme ještě, že matice $(A|B) \in \mathbb{M}_{n,n+1}$, která vznikne připojením sloupce B k matici A , se nazývá *rozšířená matice soustavy*.

Klasifikace metod

Metody numerického řešení systému (3.1) rozdělíme do dvou skupin.

(A) Přímé/finitní metody

Po konečném počtu kroků dostaneme (při počítání bez zaokrouhlování) přesné řešení. Jelikož zaokrouhlování se nevyhneme, získané řešení bude pouze přibližné.

(B) Iterační metody

Konstruujeme (obecně nekonečnou) posloupnost vektorů $\mathbf{x}^{(0)}$, $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, \dots , která konverguje k přesnému řešení. Po konečném počtu kroků, když je splněna vhodná zastavovací podmínka, výpočet zastavíme. Nalezneme tedy jen přibližné řešení.

3.1 Přímé metody

Základem většiny těchto metod je Gaussova¹ eliminační metoda (GEM), se kterou se studenti seznamují v úvodním kurzu lineární algebry. Připomeneme příslušný algoritmus a všimneme si, jaká úskalí přináší při výpočtech na počítači, kdy dochází k zaokrouhlování.

3.1.1 Gaussova eliminační metoda

Při této metodě nahrazujeme daný systém postupně ekvivalentními systémy, tj. systémy které mají totéž řešení. Na konci tohoto řetězce dostaneme systém, který lze snadno řešit. Při vytváření ekvivalentních systémů používáme tzv. *elementární řádkové úpravy matice*. Ty jsou tří typů:

- 1) výměna dvou řádků matice,
- 2) vynásobení řádku matice libovolným nenulovým číslem,
- 3) přičtení násobku jednoho řádku matice k *jinému* řádku matice.

¹**Johann Carl Friedrich Gauss** (1777–1855) — vynikající německý matematik, astronom a kartograf. Jeden z největších matematiků všech dob. Jeho práce podstatně ovlivnily mnoho matematických oborů.

Provedeme-li posloupnost libovolných elementárních řádkových úprav s rozšířenou maticí $(A|B)$ soustavy (3.2), dostaneme matici $(U|D)$, kde U je opět čtvercová matice a D je sloupcová matice. Ze základního kurzu lineární algebry je známo, že systém $U\mathbf{x} = D$ je s výchozím systémem $A\mathbf{x} = B$ ekvivalentní. Cílem je provést úpravy tak, aby matice U byla horní trojúhelníková, protože pak je systém $U\mathbf{x} = D$ snadno řešitelný.

Zmíněný algoritmus nyní popíšeme přesněji. GEM sestává ze dvou částí: *přímého chodu*, v němž se vytvoří ekvivalentní systém s horní trojúhelníkovou maticí soustavy, a *zpětného chodu*, v němž se tento speciální systém vyřeší.

Přímý chod GEM

Označme $A^{(0)} = A$, $B^{(0)} = B$. Nyní postupně vytvoříme ekvivalentní systémy $A^{(0)}\mathbf{x} = B^{(0)}$, $A^{(1)}\mathbf{x} = B^{(1)}$, \dots , $A^{(n-1)}\mathbf{x} = B^{(n-1)}$ takové, že v matici $A^{(k)}$, $k = 1, \dots, n-1$, budou všechny prvky pod hlavní diagonálou v prvních k sloupcích nulové. Tedy matice $A^{(n-1)}$ bude horní trojúhelníková.

První krok: Vytvoření systému $A^{(1)}\mathbf{x} = B^{(1)}$

(1) Položíme $A^{(1)} = A^{(0)}$ a $B^{(1)} = B^{(0)}$.

(2) Předpokládejme, že prvek $a_{11}^{(0)} \neq 0$. Nazýváme ho *hlavní prvek* nebo *pivot*.

První řádky nových matic už nebudeme měnit. Ostatní budeme modifikovat.

Uurčíme, kolikanásobky prvního řádku musíme postupně odečíst od druhého až posledního řádku, abychom v prvním sloupci dostali pod hlavní diagonálou nuly. Označme tato čísla l_{i1} , $i = 2, \dots, n$. Tedy

$$l_{21} = a_{21}^{(1)} / a_{11}^{(1)}, \quad l_{31} = a_{31}^{(1)} / a_{11}^{(1)}, \quad \dots, \quad l_{n1} = a_{n1}^{(1)} / a_{11}^{(1)}.$$

Nyní postupně odečteme od druhého řádku l_{21} -násobek prvního řádku, od třetího

řádku l_{31} -násobek prvního řádku atd. Nové prvky tedy budou (symbol $:=$ značí, že prvky v matici přepíšeme; napravo je použita původní hodnota, levá strana pak značí novou hodnotu):

$$a_{ij}^{(1)} := a_{ij}^{(1)} - l_{i1}a_{1j}^{(1)}, \quad i = 2, \dots, n, \quad j = 2, \dots, n.$$

Totéž uděláme s vektorem $B^{(1)}$. Jeho nové prvky budou:

$$b_i^{(1)} := b_i^{(1)} - l_{i1}b_1^{(1)}, \quad i = 2, \dots, n.$$

Dostaneme tudíž matice

$$A^{(1)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{pmatrix}, \quad B^{(1)} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ \vdots \\ b_n^{(1)} \end{pmatrix}.$$

Matice $A^{(1)}$ má v prvním sloupci pod hlavní diagonálou nuly.

Obecný k -tý krok, $k \leq n - 1$: Vytvoření systému $A^{(k)}x = B^{(k)}$

(1) Položíme $A^{(k)} = A^{(k-1)}$ a $B^{(k)} = B^{(k-1)}$.

(2) Protože matice $A^{(k-1)}$ měla v prvních $k - 1$ sloupcích pod hlavní diagonálou nuly, má tuto vlastnost i matice $A^{(k)}$.

Předpokládejme, že prvek $a_{kk}^{(k)} \neq 0$. Nazýváme ho *hlavní prvek* nebo *pivot*.

Prvních k řádků nových matic už nebudeme měnit. Ostatní budeme modifikovat.

Určíme, kolikanásobky k -tého řádku musíme postupně odečíst od $(k + 1)$ -ního až posledního řádku, abychom v k -tém sloupci dostali pod hlavní diagonálou nuly.

Označme tato čísla l_{ik} , $i = k + 1, \dots, n$. Tedy

$$l_{k+1,k} = a_{k+1,k}^{(k)} / a_{kk}^{(k)}, \quad l_{k+2,k} = a_{k+2,k}^{(k)} / a_{kk}^{(k)}, \quad \dots, \quad l_{nk} = a_{nk}^{(k)} / a_{kk}^{(k)}.$$

Nyní postupně odečteme od $(k + 1)$ -ního řádku $l_{k+1,k}$ -násobek k -tého řádku, od $(k + 2)$ -hého řádku $l_{k+2,k}$ -násobek k -tého řádku atd. Nové prvky tedy budou:

$$a_{ij}^{(k)} := a_{ij}^{(k-1)} - l_{ik} a_{kj}^{(k-1)}, \quad i = k + 1, \dots, n, \quad j = k + 1, \dots, n.$$

Totéž uděláme s vektorem $B^{(k)}$. Jeho nové prvky budou:

$$b_i^{(k)} := b_i^{(k-1)} - l_{ik} b_k^{(k-1)}, \quad i = k + 1, \dots, n.$$

Dostaneme tudíž matice

$$A^{(k)} = \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \cdots & a_{1,k-1}^{(k)} & a_{1k}^{(k)} & a_{1,k+1}^{(k)} & \cdots & a_{1n}^{(k)} \\ 0 & a_{22}^{(k)} & \cdots & a_{2,k-1}^{(k)} & a_{2k}^{(k)} & a_{2,k+1}^{(k)} & \cdots & a_{2n}^{(k)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{kk}^{(k)} & a_{k,k+1}^{(k)} & \cdots & a_{kn}^{(k)} \\ 0 & 0 & \cdots & 0 & 0 & a_{k+1,k+1}^{(k)} & \cdots & a_{k+1,n}^{(k)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & a_{n,k+1}^{(k)} & \cdots & a_{nn}^{(k)} \end{pmatrix},$$

$$B^{(k)} = \begin{pmatrix} b_1^{(k)} \\ b_2^{(k)} \\ \vdots \\ b_n^{(k)} \end{pmatrix}.$$

Matice $A^{(k)}$ má v prvních k sloupcích pod hlavní diagonálou nuly.

Označíme $U = A^{(n-1)}$ a $D = B^{(n-1)}$. Matice U je horní trojúhelníková s nenulovými prvky na diagonále, takže je regulární. Zavedeme ještě dolní trojúhelníkovou matici L , jež má na hlavní diagonále jedničky a pod ní prvky l_{ij} , $i = 2, \dots, n$, $j = 1$ až $i - 1$, které jsme získali v předchozím postupu. Význam této matice uvidíme později. Tedy

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 & \dots & 0 \\ l_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & 1 \end{pmatrix}. \quad (3.3)$$

Věta 3.1 *Soustavu $Ax = B$ lze řešit GEM bez výběru hlavních prvků právě tehdy, když všechny horní rohové hlavní minory matice A jsou nenulové, tj. když*

$$a_{11} \neq 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0, \dots, \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} \neq 0.$$

GEM s částečným výběrem hlavních prvků

Pokud v algoritmu GEM při provádění k -tého kroku, $1 \leq k \leq n - 1$, dojde k tomu, že hlavní prvek $a_{kk}^{(k)} = 0$, výpočet zhavaruje. Vzhledem k tomu, že matice $A^{(k-1)}$ má plnou hodnost, tj. $h(A^{(k-1)}) = n$, musí být některý z poddiagonálních prvků k -tého sloupce této matice nenulový, tj. $a_{rk}^{(k-1)} \neq 0$ pro některé r , $k + 1 \leq r \leq n$. Pak můžeme v soustavě $A^{(k-1)}\mathbf{x} = B^{(k-1)}$ vyměnit k -tou a r -tou rovnici a pokračovat standardním krokem GEM.

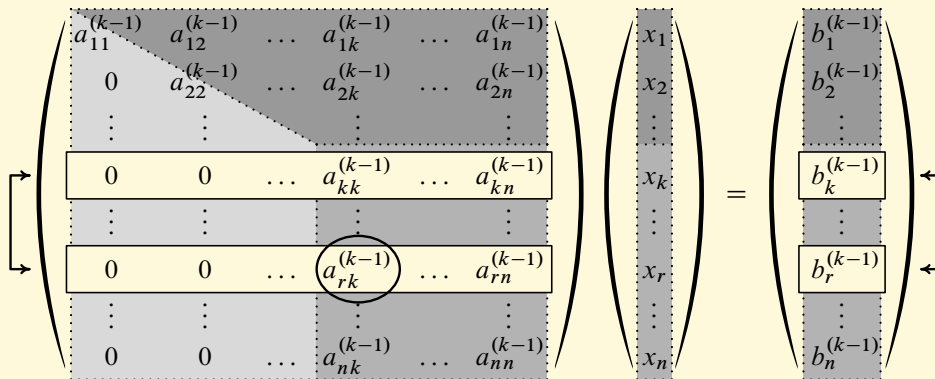
Vzhledem k tomu, že však předpokládáme, že při výpočtech dochází k zaokrouhlovacím chybám, problém může nastat, i když je hlavní prvek sice nenulový, ale hodně malý ve

srovnání s dosud neeliminovanými pod ním ležícími prvky k -tého sloupce. Koeficienty l_{ik} , $1 \leq k \leq n-1$, $k+1 \leq i \leq n$, mohou být obrovské, což má za následek, že v dalších výpočtech dojde k velkým zaokrouhlovacím chybám a získané výsledné hodnoty pro neznámé x_1, \dots, x_n jsou naprosto chybné.

Abychom se právě popsané situaci vyhnuli, provádíme na začátku každého kroku GEM tzv. *částečný výběr hlavního prvku (pivota)*. Z prvků $a_{ik}^{(k-1)}$, kde $i = k, \dots, n$, vybereme ten, který je v absolutní hodnotě největší. Nechť

$$|a_{rk}^{(k-1)}| = \max_{k \leq i \leq n} |a_{ik}^{(k-1)}|.$$

Pak vyměníme k -tou a r -tou rovnicí v soustavě $A^{(k-1)}\mathbf{x} = B^{(k-1)}$. Formálně tuto výměnu provedeme následovně. Označme $P^{(k)}$ tzv. *permutační matici*, která vznikne z jednotkové matice E záměnou k -tého a r -tého řádku. Z definice násobení matic je zřejmé, že vynásobením zleva touto maticí má za následek záměnu zmíněných řádků. Položíme tudíž $A^{(k)} = P^{(k)}A^{(k-1)}$ a $B^{(k)} = P^{(k)}B^{(k-1)}$. Dále pokračujeme tak, jak v GEM bez výběru hlavních prvků. Výměna rovnic před zahájením k -tého kroku je znázorněna na obr. 3.1.



Obr. 3.1: GEM s částečným výběrem hlavních prvků.

Prvek, který se stane hlavním prvkem, je zakroužkovan.

GEM s úplným výběrem hlavních prvků

Někdy se stává, že při použití GEM nestačí ani částečný výběr hlavních prvků a koeficienty l_{ij} je třeba zmenšit ještě více. Pak se používá tzv. *úplný výběr hlavních prvků (pivotů)*. Tentokrát se v k -tém kroku, kde $1 \leq k \leq n - 1$, vybere z celé dosud neeliminované části matice $A^{(k-1)}$ prvek, který je v absolutní hodnotě největší. Nechť

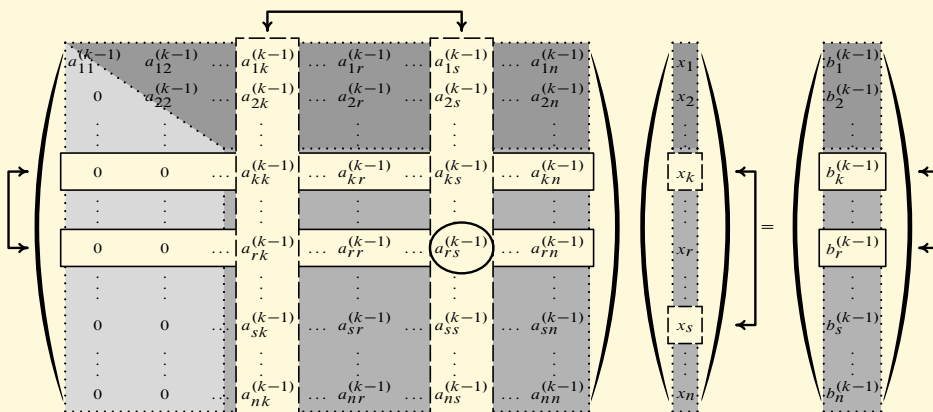
$$|a_{rs}^{(k-1)}| = \max_{\substack{k \leq i \leq n \\ k \leq j \leq n}} |a_{ij}^{(k-1)}|.$$

Pak vyměníme k -tou a r -tou rovnici v soustavě $A^{(k-1)}\mathbf{x} = B^{(k-1)}$. Dále vyměníme v matici $A^{(k-1)}$ k -tý a s -tý sloupec, což odpovídá změně pořadí neznámých.

Formálně tyto výměny provedeme následovně. Označme $P_L^{(k)}$ permutační matici, která vznikne z jednotkové matice E záměnou k -tého a r -tého řádku a $P_R^{(k)}$ permutační matici, která vznikne z jednotkové matice E záměnou k -tého a s -tého sloupce (nebo řádku, výsledek je stejný). Výměna potřebných řádků resp. sloupců se docílí násobením zleva resp. zprava těmito maticemi. Položíme tudíž $A^{(k)} = P_L^{(k)} A^{(k-1)} P_R^{(k)}$ a $B^{(k)} = P_L^{(k)} B^{(k-1)}$. Navíc označíme $\mathbf{x}^{(0)} = \mathbf{x}$ (vektor neznámých s jejich výchozím pořadím)

a $\mathbf{x}^{(k)} = P_R^{(k)} \mathbf{x}^{(k-1)}$ (vektor neznámých s jejich pořadím po k -tém kroku). Řešit tudíž budeme vlastně systém $A^{(k)} \mathbf{x}^{(k)} = B^{(k)}$. Dále pokračujeme tak, jak v GEM bez výběru hlavních prvků. Výměna rovnic a pořadí neznámých před zahájením k -tého kroku je pro $r < s$ znázorněna na obr. 3.2 (může být i $r > s$ nebo $r = s$).

Změna pořadí neznámých se obvykle zaznamenává v pomocném vektoru \mathbf{p} , jehož složky jsou na počátku čísla 1 až n . V každém kroku se s výměnou sloupců vymění i odpovídající složky vektoru \mathbf{p} , takže na konci přímého chodu GEM složky tohoto vektoru, jež jsou permutací čísel 1 až n , udávají správné pořadí neznámých. Z nich se vytvoří permutační matice $P_R = P_R^{(n-1)} \cdots P_R^{(2)} P_R^{(1)}$. Platí $\mathbf{x}^{(n-1)} = P_R \mathbf{x}^{(0)}$.



Obr. 3.2: GEM s úplným výběrem hlavních prvků.

Prvek, který se stane hlavním prvkem, je zakroužkován.

Výpočtová náročnost GEM

Všimneme si, kolik aritmetických operací je třeba k vyřešení soustavy lineárních rovnic $Ax = B$ pomocí GEM. Vzhledem k tomu, jak dlouho trvají jednotlivé operace, je vhodné rozdělit je do dvou skupin: *sčítací* (sčítání a odčítání) a *násobící* (násobení a dělení). Tabulka 3.1 uvádí příslušné počty (odvození viz [22, str. 100]).

Hodnoty jsou polynomy v proměnné n , která udává počet neznámých. Protože pro velká n je rozhodující nejvyšší mocnina, můžeme pomocí symbolu O (definice 1.2) pro $n \rightarrow \infty$ zapsat, že počet sčítacích i násobících operací přímého chodu GEM je

	Přímý chod	Zpětný chod	Celkově
Sčítací operace:	$\frac{n^3 - n}{3}$	$\frac{n^2 - n}{2}$	$\frac{2n^3 + 3n^2 - 5n}{6}$
Násobící operace:	$\frac{2n^3 + 3n^2 - 5n}{6}$	$\frac{n^2 + n}{2}$	$\frac{n^3 + 3n^2 - n}{3}$

Tab. 3.1: Počty operací v GEM

$\frac{n^3}{3} + O(n^2)$, kdežto u zpětného chodu je to pouze $\frac{n^2}{2} + O(n)$. Přímý chod je tedy co do počtu operací daleko náročnější.

Navíc u GEM s částečným nebo úplným výběrem hlavních prvků je třeba při hledání největšího prvku provést srovnání čísel. Jejich počty jsou (viz [22, str. 102]):

$$\text{Částečný výběr: } \frac{n^2 - n - 2}{2} = \frac{n^2}{2} + O(n)$$

$$\text{Úplný výběr: } \frac{2n^3 + 3n^2 - 5n}{6} = \frac{n^3}{3} + O(n^2)$$

Obvykle používáme GEM s částečným výběrem hlavních prvků. Praktické i teoretické poznatky totiž ukazují, že je postačující k tomu, aby zaokrouhlovací chyby zůstaly malé a nezneškodily výsledky (viz str. 250). Navíc počet operací je daleko nižší než u GEM s úplným výběrem hlavních prvků.

Speciální matice

V řadě aplikací vzniká ve finální fázi potřeba numericky řešit systémy lineárních rovnic, jejichž matice soustavy mají speciální vlastnosti. Díky tomu se některé postupy zjednoduší,

některé předpoklady jsou automaticky splněny apod. Uvedeme nyní několik takových nejčastějších typů matic.

Definice 3.2 Čtvercová matice $A \in \mathbb{M}_n$ se nazývá *ryze řádkově diagonálně dominantní*, jestliže platí

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, \dots, n.$$

Čtvercová matice $A \in \mathbb{M}_n$ se nazývá *ryze sloupcově diagonálně dominantní*, jestliže platí

$$|a_{jj}| > \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}|, \quad j = 1, \dots, n.$$

Tedy absolutní hodnota libovolného diagonálního prvku je větší než součet absolutních hodnot zbývajících prvků v témže řádku (sloupci). Zřejmě všechny diagonální prvky takové matice musí být nenulové.

Věta 3.3 *Ryze řádkově nebo sloupcově diagonálně dominantní matice je regulární.*

Důkaz. Pripusťme, že A je ryze řádkově diagonálně dominantní a je singulární. Pak má homogenní soustava $A\mathbf{x} = \mathbf{0}$ netriviální řešení \mathbf{x} . Nechť

$$|x_r| = \max\{|x_1|, |x_2|, \dots, |x_n|\}.$$

Tedy x_r je složka, která je v absolutní hodnotě největší. Zřejmě $|x_r| > 0$.

Uvažujme nyní r -tou rovnicí soustavy $A\mathbf{x} = \mathbf{0}$. Osamostatníme z ní x_r .

$$a_{rr}x_r = -a_{r1}x_1 - \dots - a_{r,r-1}x_{r-1} - a_{r,r+1}x_{r+1} - \dots - a_{rn}x_n.$$

Vypočítáme absolutní hodnotu obou stran předchozí rovnosti. S využitím pravidel pro počítání s absolutní hodnotou (viz str. 40) vyjde

$$\begin{aligned} |a_{rr}| \cdot |x_r| &= |-a_{r1}x_1 - \dots - a_{r,r-1}x_{r-1} - a_{r,r+1}x_{r+1} - \dots - a_{rn}x_n| \leq \\ &\leq |a_{r1}| \cdot |x_1| + \dots + |a_{r,r-1}| \cdot |x_{r-1}| + \\ &\quad + |a_{r,r+1}| \cdot |x_{r+1}| + \dots + |a_{rn}| \cdot |x_n| \leq \\ &\leq (|a_{r1}| + \dots + |a_{r,r-1}| + |a_{r,r+1}| + \dots + |a_{rn}|) |x_r|. \end{aligned}$$

Po vydělení kladným číslem $|x_r|$ dostaneme

$$|a_{rr}| \leq |a_{r1}| + \cdots + |a_{r,r-1}| + |a_{r,r+1}| + \cdots + |a_{rn}|,$$

což je spor s předpokladem, že matice A je ryze řádkově diagonálně dominantní.

Je-li A ryze sloupcově diagonálně dominantní, je podle první části důkazu matice A^T regulární, protože A^T je ryze řádkově diagonálně dominantní. Protože A a A^T mají stejný determinant, je A regulární. \square

Další vlastnost je definovaná pro symetrické matice, tj. takové čtvercové matice, pro něž platí $A^T = A$. S následujícím důležitým pojmem jsme se již setkali na str. 60.

Definice 3.4 Symetrická matice $A \in \mathbb{M}_n$ se nazývá *pozitivně definitní*, jestliže pro libovolný nenulový vektor $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$, platí, že $\mathbf{x}^T A \mathbf{x} > 0$.

Přímo z definice snadno plynou některé vlastnosti pozitivně definitních matic. Taková matice je nutně regulární. Jinak by existoval $\mathbf{x} \neq \mathbf{0}$ takový, že $A \mathbf{x} = \mathbf{0}$. Pak by ovšem pro tento vektor platilo $\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T \mathbf{0} = 0$, což je spor. Dále pro vektor, jehož i -tá složka je

jednička a ostatní jsou nulové, tj. $\mathbf{x}^T = [0, \dots, 0, 1, 0, \dots, 0]$, vyjde $\mathbf{x}^T A \mathbf{x} = a_{ii}$. Tedy všechny diagonální prvky musí být kladné.

Existuje nutná a postačující podmínka pozitivní definitnosti. Její použití v případě rozsáhlé matice ovšem není snadné.

Věta 3.5 (Sylvestrov¹ kritérium) *Symetrická matice $A \in \mathbb{M}_n$ je pozitivně definitní právě tehdy, když všechny horní rohové hlavní minory matice A jsou kladné, tj. když*

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \dots, \quad \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} > 0.$$

Důkaz viz [18, str. 56] nebo [52, str. 204 a 283].

Poznámka 3.6 Je-li $A \in \mathbb{M}_n$ symetrická matice a pro libovolný nenulový vektor $\mathbf{x} \in \mathbb{R}^n$ platí $\mathbf{x}^T A \mathbf{x} \geq 0$, nazývá se A *pozitivně semidefinitní*. Pozitivně semidefinitní matice, která není pozitivně definitní, je určitě singulární.

¹**James Joseph Sylvester** (1814–1897) — anglický matematik. Zabýval se algebrou, maticemi a teoretickou a aplikovanou kinematikou.

Lze ukázat, že A je pozitivně semidefinitní, právě když všechny hlavní minory (tj. determinanty podmatic tvořených týmiž řádky a sloupci) jsou nezáporné.

Obdobně se zavádí pojmy *negativně definitní* ($\mathbf{x}^T A \mathbf{x} < 0$ pro $\mathbf{x} \neq \mathbf{0}$) a *negativně semidefinitní* ($\mathbf{x}^T A \mathbf{x} \leq 0$ pro libovolné \mathbf{x}) matice. Snadno je vidět, že A je negativně (semi)definitní, právě když je $-A$ pozitivně (semi)definitní. Odtud lze snadno odvodit kritéria ověřující tyto vlastnosti pomocí hlavních minorů.

Symetrická matice, která nemá žádnou ze čtyř zmíněných vlastností, se nazývá *indefinitní*.

Symetrické matice mají řadu důležitých vlastností, viz např. [18, str. 44–65] a hrají významnou roli v mnoha částech matematiky včetně numerických metod.

Nás zajímají tyto speciální matice v souvislosti s GEM. Zřejmě každá horní rohová podmatice ryze řádkově nebo sloupcově diagonálně dominantní matice má tutéž vlastnost. Z vět 3.1, 3.3 a 3.5 proto plyne následující tvrzení.

Věta 3.7 *Je-li matice $A \in \mathbb{M}_n$ ryze řádkově nebo sloupcově diagonálně dominantní nebo symetrická pozitivně definitní, lze soustavu $A\mathbf{x} = B$ s libovolnou pravou stranou $B \in \mathbb{R}^n$ řešit GEM bez výběru hlavních prvků.*

Často se setkáváme s (velkými) maticemi, které mají většinu prvků nulových. Takovým maticím se říká *řídke*. Tato vlastnost je důležitá z hlediska paměťových nároků, nulové

prvky není nutné ukládat. Obvyklé jsou případy, kdy nenulové prvky jsou pouze na hlavní diagonále a v její blízkosti.

Definice 3.8 Matice $A \in \mathbb{M}_n$ se nazývá *pásová*, jestliže existují čísla $p, q \in \mathbb{N}_0$, $0 \leq p, q \leq n - 1$, taková, že pro $j - i > p$ nebo $i - j > q$ platí $a_{ij} = 0$. Číslo $p + q + 1$ je šířka pásu.

Předchozí vlastnost znamená, že A může mít nenulové prvky pouze na hlavní diagonále, p naddiagonálách a q poddiagonálách. Časté jsou zejména třídiagonální matice ($p = q = 1$, tj. šířka pásu je tři). Pro $p = q = 0$ jde o diagonální matici.

Je-li matice A pásová, je celkem zřejmé, že v přímém chodu GEM bez výběru hlavních prvků nemohou vzniknout v maticích $A^{(k)}$, $k = 1, \dots, n - 1$, nenulové prvky mimo tento pás. Je-li pás úzký, počet operací potřebných k eliminaci se významně redukuje a i paměťové nároky na uložení průběžných výsledků se znatelně sníží.

3.1.2 Metoda LU rozkladu

Definice 3.9 Jestliže lze matici $A \in \mathbb{M}_n$ vyjádřit ve tvaru součinu $A = LU$, kde L je dolní trojúhelníková matice a U je horní trojúhelníková matice, nazývá se toto vyjádření *LU rozklad matice A* .

Nás bude zajímat případ, kdy matice A je regulární. Pak i matice L a U musí být regulární (součin dvou čtvercových matic je regulární, právě když jsou oba činitele regulární). Protože jsou trojúhelníkové, musí být tudíž všechny jejich diagonální prvky nenulové.

V přímém chodu GEM bez výběru hlavních prvků jsme z matice A získali dvě matice — horní trojúhelníkovou matici U a dolní trojúhelníkovou matici L , viz (3.3). Matice U byla dále využita ve zpětném chodu GEM. Velmi důležitá je však i matice L . Lze dokázat (viz např. [22, str. 98]), že tyto matice určují LU rozklad matice A , tj.

$$A = LU. \quad (3.5)$$

Všimneme si existence takového rozkladu. Jestliže je možné provést přímý chod GEM bez výběru hlavních prvků, zmíněný rozklad existuje.

Nechť naopak LU rozklad existuje. Označme A_k , L_k a U_k podmatice matic A , L a U tvořené prvními k řádky a prvními k sloupci, $k = 1, \dots, n$. Zřejmě matice L_k a U_k budou trojúhelníkové a regulární (protože mají nenulové diagonální prvky). Snadno je vidět, že $A_k = L_k U_k$, takže A_k jsou regulární. To znamená, že všechny horní rohové hlavní minory jsou nenulové, takže podle věty 3.1 je možné provést GEM bez výběru hlavních prvků. Dostáváme následující výsledek:

Věta 3.10 *Nechť $A \in \mathbb{M}_n$ je regulární matice. Pak LU rozklad této matice existuje právě tehdy, když lze systém $Ax = B$, kde $B \in \mathbb{R}^n$ je libovolný sloupec, řešit GEM bez výběru hlavních prvků.*

Připomeňme, že podle věty 3.1 je další ekvivalentní podmínkou existence LU rozkladu nenulovost všech horních rohových hlavních minorů matice A .

Poznámka 3.11 LU rozklad regulární matice A není určen jednoznačně. Abychom mohli tuto otázku vyřešit, budeme potřebovat některé poznatky o trojúhelníkových maticích, které lze snadno ověřit. O všech maticích předpokládáme, že jsou čtvercové řádu n .

- Součin dvou horních (dolních) trojúhelníkových matic je opět horní (dolní) trojúhelníková matice.
- Inverzní matice k regulární horní (dolní) trojúhelníkové matici je opět horní (dolní) trojúhelníková matice.
- Jsou-li R a S horní (dolní) trojúhelníkové matice a $T = RS$, pak $t_{ii} = r_{ii}s_{ii}$, $i = 1$ až n . Tedy diagonální prvky součinu dvou horních (dolních) trojúhelníkových matic jsou rovny součinu diagonálních prvků jednotlivých činitelů.
- Je-li R regulární trojúhelníková matice, jsou diagonální prvky matice R^{-1} rovny $1/r_{ii}$, $i = 1$ až n . Tedy diagonální prvky matice inverzní k trojúhelníkové matici jsou rovny převráceným hodnotám diagonálních prvků původní matice.

Všechna předchozí tvrzení zejména platí pro diagonální matice, které jsou současně horní i dolní trojúhelníkové.

Ukazuje se, že rozklad $A = LU$ je jednoznačně určen, jestliže předepíšeme všechny diagonální prvky buď matice L nebo U . Předepsané hodnoty musí být samozřejmě nenulové.

Předpokládejme, že $A = LU$ je nějaký LU rozklad matice A a zvolme nenulová čísla p_1 až p_n . Ukážeme, že pak existuje LU rozklad, kde např. první činitel má na diagonále předepsaná

čísla.

Označme D diagonální matici, jejíž diagonální prvky jsou čísla p_i/l_{ii} . Pak platí

$$A = LU = LDD^{-1}U = L_1U_1, \quad \text{kde } L_1 = LD, \quad U_1 = D^{-1}U.$$

Matice L_1 je dolní trojúhelníková a U_1 je horní trojúhelníková. Přitom L_1 má na diagonále čísla p_1, \dots, p_n .

Připusťme, že existuje ještě druhý LU rozklad $A = L_2U_2$, kde matice L_2 má na diagonále tatáž čísla p_1, \dots, p_n . Pak

$$L_1U_1 = L_2U_2, \quad \text{odkud} \quad L_1^{-1}L_2 = U_1U_2^{-1}.$$

Matice $L_1^{-1}L_2$ je dolní trojúhelníková a matice $U_1U_2^{-1}$ je horní trojúhelníková. Protože se rovnají, musí být rovny nějaké diagonální matici D . Jelikož L_1 a L_2 mají stejné diagonální prvky, musí být D jednotková matice. Z rovností $L_1^{-1}L_2 = E$ a $U_1U_2^{-1} = E$ plyne, že $L_1 = L_2$ a $U_1 = U_2$, takže LU rozklad je určen jednoznačně.

Použití LU rozkladu k řešení lineárních systémů

Předpokládejme nyní, že jsme pomocí přímého chodu GEM našli LU rozklad matice soustavy A ve tvaru $A = LU$. Ukážeme, jak jej můžeme použít k řešení soustavy $Ax = B$ s libovolnou pravou stranou B . Tento postup se nazývá *metoda LU rozkladu*.

Soustava má tvar $LUx = B$. Označíme-li $Ux = y$, je vektor y řešením soustavy $Ly = B$. Ta má tvar (připomeňme, že L má na hlavní diagonále jedničky — viz (3.3))

$$\begin{aligned}
 y_1 &= b_1, \\
 l_{21}y_1 + y_2 &= b_2, \\
 \dots & \\
 l_{n-1,1}y_1 + l_{n-1,2}y_2 + \dots + y_{n-1} &= b_{n-1}, \\
 l_{n,1}y_1 + l_{n,2}y_2 + \dots + l_{n,n-1}y_{n-1} + y_n &= b_n.
 \end{aligned} \tag{3.6}$$

Tu snadno vyřešíme. Z první rovnice máme $y_1 = b_1$. Po dosazení do druhé rovnice vypočítáme $y_2 = b_2 - l_{21}y_1$. A tak dále až z poslední rovnice vypočítáme y_n .

Se soustavou $Ux = y$ jsme se setkali ve zpětném chodu GEM — viz (3.4). Snadno ji vyřešíme, tentokrát od konce.

Použití metody LU rozkladu je zejména výhodné, když máme řešit více soustav $Ax = B$ s různými pravými stranami B , ale touž maticí A . Stačí jednou najít LU rozklad matice A , což je výpočtově mnohem náročnější část, a pak jen pro každou pravou stranu B vyřešit dvojici trojúhelníkových soustav, což je daleko méně náročné.

Všimneme si výpočtové náročnosti metody LU rozkladu. Nejprve musíme najít rozklad $A = LU$, což je rovnocenné přímému chodu GEM. Neupravujeme však žádný sloupec B , takže počet sčítacích i násobících operací se oproti tabulce 3.1 sníží o $\frac{n^2-n}{2}$. Nicméně asymptoticky tyto počty zůstanou stejné, tj. $\frac{n^3}{3} + O(n^2)$. Pak provedeme zpětný chod metody LU rozkladu, tedy vyřešíme dvě trojúhelníkové soustavy, což odpovídá dvěma zpětným chodům GEM, tj. $n^2 + O(n)$ sčítacím i násobícím operacím.

Použití metody si ukážeme na příkladu. Protože nám jde o pochopení algoritmu a ne o problémy se zaokrouhlováním, výpočty provedeme v celočíselné aritmetice.

Příklad 3.12 Metodou LU rozkladu najděte řešení soustavy

$$\begin{aligned}x_1 + 2x_2 + x_3 &= 3, \\x_1 + x_2 - 2x_3 &= 5, \\3x_1 - x_2 - x_3 &= -2.\end{aligned}$$

Řešení. Nejprve ověříme, že metodu LU rozkladu lze použít. Podle vět 3.10 a 3.1 musí být všechny horní rohové hlavní minory nenulové. Dostaneme:

$$1 \neq 0, \quad \begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} = -1 \neq 0, \quad \begin{vmatrix} 1 & 2 & 1 \\ 1 & 1 & -2 \\ 3 & -1 & -1 \end{vmatrix} = -17 \neq 0.$$

LU rozklad matice soustavy A tedy existuje. V následujících úpravách čísla v kulatých závorkách značí řádky, koeficienty u nich udávají, čím řádky násobíme. Čísla řádků se vždy vztahují k předchozí matici, výsledky se zapisují do toho řádku nové matice, vedle kterého údaje stojí.

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & -2 \\ 3 & -1 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & -3 \\ 0 & -7 & -4 \end{pmatrix} \begin{matrix} (2) - 1(1) \\ (3) - 3(1) \end{matrix} \sim \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & -3 \\ 0 & 0 & 17 \end{pmatrix} \begin{matrix} \\ (3) - 7(2) \end{matrix}$$

Koeficienty l_{ij} udávají, kolikanásobek j -tého řádku odečítáme od i -tého řádku — viz algoritmus přímého chodu GEM. Tedy $l_{21} = 1$, $l_{31} = 3$ a $l_{32} = 7$. Nyní již můžeme

sestavit matice L (na hlavní diagonále má jedničky) a U — viz (3.3).

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 7 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & -3 \\ 0 & 0 & 17 \end{pmatrix}.$$

Řešení soustavy $A\mathbf{x} = B$, tj. $LU\mathbf{x} = B$, dostaneme postupným řešením soustav $L\mathbf{y} = B$ a $U\mathbf{x} = \mathbf{y}$. Vyjde:

$$\begin{array}{rcl} y_1 & = & 3, \\ y_1 + y_2 & = & 5, \\ 3y_1 + 7y_2 + y_3 & = & -2, \end{array} \quad \text{odkud} \quad \begin{array}{l} y_1 = 3, \\ y_2 = 5 - y_1 = 2, \\ y_3 = -2 - 3y_1 - 7y_2 = -25, \end{array}$$

$$\begin{array}{rcl} x_1 + 2x_2 + x_3 & = & 3, \\ -x_2 - 3x_3 & = & 2, \\ 17x_3 & = & -25, \end{array} \quad \text{odkud} \quad \begin{array}{l} x_3 = -\frac{25}{17}, \\ x_2 = -2 - 3x_3 = \frac{41}{17}, \\ x_1 = 3 - 2x_2 - x_3 = -\frac{6}{17}. \end{array} \quad \blacktriangle$$

LU rozklad s částečným výběrem hlavních prvků

Jak víme, ne každá regulární matice A má LU rozklad. Ten podle vět 3.10 a 3.1 existuje, právě když jsou všechny horní rohové hlavní minory nenulové. Neexistence LU rozkladu se v přímém chodu GEM projeví tím, že v některém kroku je hlavní prvek $a_{kk}^{(k)} = 0$. Pak je nutné sáhnout k výměně řádků. Jak jsme ale vysvětlili v GEM s částečným výběrem hlavních prvků, problémem jsou i malé hlavní prvky $a_{kk}^{(k)}$, které způsobují velké zaokrouhlovací chyby.

Stejný problém se projevuje i při LU rozkladu. Abychom se těmto potížím vyhnuli, i v tomto případě používáme částečný výběr hlavních prvků. Postup je obdobný jako u GEM — viz str. 215. Je však třeba rozmyslet si, jak získáme dolní trojúhelníkovou matici L a horní trojúhelníkovou matici U .

Při GEM s částečným výběrem hlavních prvků provádíme (pokud je to nutné) výměny řádků upravované matice A . Celou situaci si můžeme představit tak, že provádíme GEM bez výběru hlavních prvků, tj. bez přehazování řádků, ale ne s maticí A , avšak s jistou maticí \hat{A} , jež z A vznikne vhodným přeházením řádků. Bohužel dopředu nevíme, které řádky se mají přeházet. To zjišťujeme až během přímého chodu GEM.

Při popisu GEM s částečným výběrem hlavních prvků jsme zavedli permutační matice $P^{(k)}$, $k = 1, \dots, n - 1$. Vynásobením zleva těmito maticemi provedlo výměnu dvojice řádků. Matice $A^{(k)}$, v níž jsme v k -tém kroku eliminovali poddiagonální prvky, se proto rovnala $P^{(k)} A^{(k-1)}$. Postupné násobení zleva těmito permutačními maticemi tudíž vytvoří z matice A hledanou matici \widehat{A} . Tedy $\widehat{A} = P^{(n-1)} P^{(n-2)} \dots P^{(2)} P^{(1)} A$. Stačilo by tedy vzít matici \widehat{A} a najít její LU rozklad, tj. $\widehat{A} = LU$.

Označíme-li $P = P^{(n-1)} \dots P^{(1)}$, bude platit $PA = \widehat{A} = LU$. Matice P se rovněž nazývá *permutační*. Vznikne z jednotkové matice postupným přehazováním řádků. Tedy P má v každém řádku a každém sloupci právě jednu jedničku, její ostatní prvky jsou nulové.

Bohužel, jak již bylo řečeno, matice $P^{(1)}, \dots, P^{(n-1)}$ předem neznáme. Můžeme však vzít vektor $\mathbf{p}^T \in \mathbb{R}^n$, do něhož před zahájením výpočtu vložíme čísla 1 až n , tj. $\mathbf{p} = [1, 2, \dots, n]^T$. Když budeme v některém kroku měnit např. k -tý a r -tý řádek, vyměníme ve vektoru \mathbf{p} jeho k -tou a r -tou složku. Na konci budeme mít v \mathbf{p} poznačeno, jak se řádky matice A přeházely, a snadno vytvoříme matici P . Horní trojúhelníková matice U , kterou dostaneme, je rovněž ta, kterou potřebujeme.

Jak však dostaneme správnou dolní trojúhelníkovou matici L ? Víme, že průběžně

ukládáme po sloupcích její prvky pod hlavní diagonálou. Představme si, že bychom předem věděli, že v k -tém kroku přímého chodu GEM bude nutné vyměnit k -tý a r -tý řádek. Jaký by to mělo dopad na průběžně vytvářenou matici L , kdybychom výměnu skutečně předem provedli? Zřejmě by se vyměnil k -tý a r -tý řádek v dosud hotové části matice L , tj. prvky v prvních $k - 1$ sloupcích. Celý výpočet lze zorganizovat následovně.

- 1) Do postupně eliminovaných sloupců matice A zapisujeme pod hlavní diagonálu místo nul prvky matice L .
- 2) Pokud je třeba provést v k -tém kroku výměnu k -tého a r -tého řádku, provedeme ji s celými n -člennými řádky, tj. vyměníme vlastně jak dosud hotové části k -tého a r -tého řádku matice L (první až $(k - 1)$ -ní sloupec), tak dosud neeliminované části k -tého a r -tého řádku matice A (k -tý až n -tý sloupec).
- 3) Při následné eliminaci poddiagonálních prvků k -tého sloupce, tj. odečítání l_{ik} -násobků k -tého řádku od řádků ležících pod ním, se prvky v prvním až $(k - 1)$ -ním sloupci nemění (tvoří část matice L), do vynulovaných poddiagonálních pozic k -tého sloupce se uloží další prvky matice, tj. čísla l_{ik} , a prvky v k -tém až n -tém sloupci se standardně upraví.

Tento postup je výhodný i z hlediska úspory místa, zejména u velkých matic. Nulové poddiagonální prvky není třeba ukládat, místo nich tam uschováme prvky matice L . Taková strategie, tj. uložení prvků matice L do eliminovaných pozic pod diagonálou, je samozřejmě možná a vhodná i u LU rozkladu bez výběru hlavních prvků.

Ve výsledné matici $A^{(n-1)}$ prvky na hlavní diagonále a nad ní určují horní trojúhelníkovou matici U a prvky pod ní určují dolní trojúhelníkovou matici L (ta má na diagonále jedničky). Pomocí vektoru \mathbf{p} vytvoříme matici P . Je-li i -tý prvek \mathbf{p} roven číslu j , má matice P v i -tém řádku a j -tém sloupci jedničku, ostatní prvky i -tého řádku jsou nuly. Dále platí

$$PA = LU. \quad (3.7)$$

Mluvíme o tzv. *PLU rozkladu matice A* .

Tento rozklad můžeme využít k řešení soustavy $A\mathbf{x} = \mathbf{B}$. Po vynásobení permutační maticí P dostaneme $PA\mathbf{x} = P\mathbf{B}$, tedy $LU\mathbf{x} = P\mathbf{B}$. Označíme-li $U\mathbf{x} = \mathbf{y}$, dostáváme opět dvě trojúhelníkové soustavy $L\mathbf{y} = P\mathbf{B}$ a $U\mathbf{x} = \mathbf{y}$. Postup nazýváme *metoda LU rozkladu s částečným výběrem hlavních prvků*.

Výpočtová složitost je stejná jako u GEM s částečným výběrem hlavních prvků, tj. $\frac{n^3}{3} + O(n^2)$ sčítacích a násobících operací a $\frac{n^2}{2} + O(n)$ srovnání pro nalezení PLU

rozkladu a $n^2 + O(n)$ sčítacích a násobících operací pro řešení dvou trojúhelníkových soustav.

Poznámka 3.13 Snadno se zvaží, že při dané permutační matici P je PLU rozklad jednoznačný, když předepíšeme diagonální prvky matice L (v našem případě jsou to jedničky) nebo U , srovnejte poznámku 3.11.

Příklad 3.14 Vyřešte soustavu z příkladu 3.12 metodou LU rozkladu s částečným výběrem hlavních prvků.

Řešení. Budeme postupovat obdobně jako v příkladu 3.12. Do vynulovaných poloh pod hlavní diagonálou ale budeme psát prvky matice L (jsou v rámečcích).

$$\begin{aligned} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & -2 \\ 3 & -1 & -1 \end{pmatrix} &\sim \begin{pmatrix} 3 & -1 & -1 \\ 1 & 1 & -2 \\ 1 & 2 & 1 \end{pmatrix} \begin{matrix} (3) \\ (1) \\ (1) \end{matrix} \sim \begin{pmatrix} 3 & -1 & -1 \\ \boxed{\frac{1}{3}} & \frac{4}{3} & -\frac{5}{3} \\ \boxed{\frac{1}{3}} & \frac{7}{3} & \frac{4}{3} \end{pmatrix} \begin{matrix} (2) - \frac{1}{3}(1) \\ (3) - \frac{1}{3}(1) \end{matrix} \sim \\ &\sim \begin{pmatrix} 3 & -1 & -1 \\ \boxed{\frac{1}{3}} & \frac{7}{3} & \frac{4}{3} \\ \boxed{\frac{1}{3}} & \frac{4}{3} & -\frac{5}{3} \end{pmatrix} \begin{matrix} (3) \\ (2) \end{matrix} \sim \begin{pmatrix} 3 & -1 & -1 \\ \boxed{\frac{1}{3}} & \frac{7}{3} & \frac{4}{3} \\ \boxed{\frac{1}{3}} & \boxed{\frac{4}{7}} & -\frac{17}{7} \end{pmatrix} \begin{matrix} (3) - \frac{4}{7}(2) \end{matrix} \end{aligned}$$

Do vektoru \mathbf{p} zapíšeme výměny řádků:

$$\mathbf{p} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \rightarrow \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Nyní již můžeme sestavit matice L (na hlavní diagonále má jedničky), U a P .

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{1}{3} & \frac{4}{7} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 3 & -1 & -1 \\ 0 & \frac{7}{3} & \frac{4}{3} \\ 0 & 0 & -\frac{17}{7} \end{pmatrix}, \quad P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Ještě určíme vektor PB :

$$PB = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ -2 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \\ 5 \end{pmatrix}$$

Řešení soustavy $PA\mathbf{x} = PB$, tj. $LU\mathbf{x} = PB$, dostaneme postupným řešením soustav $L\mathbf{y} = PB$ a $U\mathbf{x} = \mathbf{y}$. Vyjde:

$$\begin{array}{rcl} y_1 & = & -2, \\ \frac{1}{3}y_1 + y_2 & = & 3, \\ \frac{1}{3}y_1 + \frac{4}{7}y_2 + y_3 & = & 5, \end{array} \quad \text{odkud} \quad \begin{array}{rcl} y_1 & = & -2, \\ y_2 & = & 3 - \frac{1}{3}y_1 = \frac{11}{3}, \\ y_3 & = & 5 - \frac{1}{3}y_1 - \frac{4}{7}y_2 = \frac{25}{7}, \end{array}$$

$$\begin{array}{rcl}
 3x_1 - x_2 - x_3 = -2, & & x_3 = -\frac{25}{17}, \\
 \frac{7}{3}x_2 + \frac{4}{3}x_3 = \frac{11}{3}, & \text{odkud} & x_2 = -2 - 3x_3 = \frac{41}{17}, \\
 -\frac{17}{7}x_3 = \frac{25}{7}, & & x_1 = 3 - 2x_2 - x_3 = -\frac{6}{17}. \quad \blacktriangle
 \end{array}$$

3.1.3 Choleského metoda

Metoda, kterou nyní uvedeme, je určena pouze pro případ, kdy matice soustavy je symetrická a pozitivně definitní. Je založena na následujícím výsledku (viz [52, str. 205]).

Věta 3.15 *Nechť A je symetrická pozitivně definitní matice. Pak existuje právě jedna dolní trojúhelníková matice L s kladnými prvky na hlavní diagonále taková, že platí $A = LL^T$.*

Vyjádření z předchozí věty se nazývá *Choleského¹ rozklad matice A* .

¹**André-Louis Cholesky** (1875–1918) (čti šolesky) — francouzský důstojník, zabýval se geodézií a kartografií. Na začátku 20. století se podílel na zeměměřických pracích na Krétě a v severní Africe. Padl krátce před koncem první světové války.

Pro výpočet prvků matice L se snadno odvodí rekurentní vzorce. Protože tato matice je dolní trojúhelníková, pro $i < k$ je $l_{ik} = 0$. Z rovnosti $A = LL^T$ dále dostaneme

$$\begin{aligned} l_{k1}^2 + l_{k2}^2 + \cdots + l_{kk}^2 &= a_{kk}, & k &= 1, \dots, n, \\ l_{k1}l_{i1} + l_{k2}l_{i2} + \cdots + l_{kk}l_{ik} &= a_{ki}, & 1 \leq k < i \leq n. \end{aligned} \quad (3.8)$$

Prvky nyní počítáme po sloupcích.

$$\begin{aligned} l_{11} &= \sqrt{a_{11}}, \\ l_{i1} &= \frac{a_{1i}}{l_{11}}, & i &= 2, \dots, n, \\ l_{kk} &= \sqrt{a_{kk} - \sum_{j=1}^{k-1} l_{kj}^2}, & k &= 2, \dots, n, \\ l_{ik} &= \frac{1}{l_{kk}} \left(a_{ki} - \sum_{j=1}^{k-1} l_{kj}l_{ij} \right), & k &= 2, \dots, n, \quad i = k+1, \dots, n. \end{aligned} \quad (3.9)$$

Při řešení soustavy $A\mathbf{x} = \mathbf{B}$ se symetrickou pozitivně definitní maticí soustavy

najdeme nejprve rozklad $A = LL^T$ a pak podobně jako u LU rozkladu řešíme dvě trojúhelníkové soustavy $Ly = B$ a $L^T x = y$. Postup nazýváme *Choleského metoda*.

Výpočtová složitost Choleského rozkladu je $\frac{n^3}{6} + O(n^2)$ sčítacích a násobících operací a výpočet n druhých odmocnin (což je zanedbatelné) a řešení dvou trojúhelníkových soustav, jak již víme, $n^2 + O(n)$ sčítacích a násobících operací.

Z první rovnosti v (3.8) je vidět, že $l_{kj} \leq \sqrt{a_{kk}}$, $k = 1, \dots, n$, $j = 1, \dots, k$, tedy prvky matice L nemohou příliš narůstat oproti diagonálním prvkům matice A .

Poznámka 3.16 Choleského rozklad existuje i pro symetrické matice, které nejsou pozitivně definitní, ale mají LU rozklad (horní rohové hlavní minory jsou nenulové, viz věty 3.10 a 3.1). Pak ovšem některé diagonální prvky matice L budou ryze imaginární (pod odmocninou ve (3.9) vyjde záporné číslo). Stane-li se to pro l_{kk} , pak i všechna l_{jk} , $j = k + 1, \dots, n$, budou ryze imaginární. Při řešení trojúhelníkových soustav se tato komplexní čísla zruší (při přesném výpočtu). Viz [22, str. 108].

Příklad 3.17 Choleského metodou najděte řešení soustavy

$$\begin{aligned}x_1 + 2x_2 - x_3 &= 3, \\2x_1 + 5x_2 - 2x_3 &= 5, \\-x_1 - 2x_2 + 3x_3 &= -2.\end{aligned}$$

Řešení. Matice soustavy A je zřejmě symetrická. Ověříme pomocí **Sylvestrova kritéria**, že je pozitivně definitní.

$$1 > 0, \quad \begin{vmatrix} 1 & 2 \\ 2 & 5 \end{vmatrix} = 1 > 0, \quad \begin{vmatrix} 1 & 2 & -1 \\ 2 & 5 & -2 \\ -1 & -2 & 3 \end{vmatrix} = 2 > 0.$$

Protože všechny horní rohové hlavní minory jsou kladné, je A skutečně pozitivně definitní.

Najdeme Choleského rozklad matice soustavy $A = LL^T$. Ze vzorců (3.9) dostaneme:

$$\begin{aligned} l_{11} &= \sqrt{a_{11}} = 1, & l_{21} &= \frac{a_{12}}{l_{11}} = 2, & l_{31} &= \frac{a_{13}}{l_{11}} = -1, \\ l_{22} &= \sqrt{a_{22} - l_{21}^2} = 1, & l_{32} &= \frac{1}{l_{22}} (a_{23} - l_{21}l_{31}) = 0, \\ l_{33} &= \sqrt{a_{33} - l_{31}^2 - l_{32}^2} = \sqrt{2}. \end{aligned}$$

Řešení soustavy $A\mathbf{x} = B$, tj. $LL^T\mathbf{x} = B$, dostaneme postupným řešením soustav

$L\mathbf{y} = B$ a $L^T\mathbf{x} = \mathbf{y}$. Vyjde:

$$\begin{array}{rcl} y_1 & = & 3, \\ 2y_1 + y_2 & = & 5, \\ -y_1 + \sqrt{2}y_3 & = & -2, \end{array} \quad \text{odkud} \quad \begin{array}{l} y_1 = 3, \\ y_2 = 5 - 2y_1 = -1, \\ y_3 = \frac{1}{\sqrt{2}}(-2 + y_1) = \frac{\sqrt{2}}{2}, \end{array}$$

$$\begin{array}{rcl} x_1 + 2x_2 - x_3 & = & 3, \\ x_2 & = & -1, \\ \sqrt{2}x_3 & = & \frac{\sqrt{2}}{2}, \end{array} \quad \text{odkud} \quad \begin{array}{l} x_3 = \frac{1}{2}, \\ x_2 = -1, \\ x_1 = 3 - 2x_2 + x_3 = \frac{11}{2}. \end{array} \quad \blacktriangle$$

Poznámka 3.18 Choleského rozklad je zřejmě speciálním případem LU rozkladu. Zatímco při hledání LU rozkladu jsme použili algoritmus GEM, pro Choleského rozklad máme vzorce (3.9) pro nalezení prvků matice L . Obdobně tzv. *přímé vzorce* lze odvodit i pro obecný případ LU rozkladu. Existují různé postupy, v jakém pořadí se počítají prvky matic L a U . U *Croutovy¹ metody* se střídavě počítají shora řádky matice U a zleva sloupce matice L , zatímco u *Banachiewiczzovy² metody* se střídavě počítají shora řádky matice U a matice L . Podrobněji viz [52, str. 197].

¹**Prescott Durand Crout** (1907–1984) (čti kraut) — americký matematik.

²**Tadeusz Banachiewicz** (1882–1954) (čti banachievič) — polský astronom a matematik.

3.1.4 Analýza chyb při numerickém řešení soustav lineárních rovnic

V předchozích oddílech jsme se zabývali přímými metodami řešení soustav lineárních rovnic $Ax = B$ s regulární maticí soustavy A . Při přesném počítání tyto metody dávají přesné řešení x . Avšak provádíme-li výpočty na počítači v číselné soustavě s plovoucí řádovou čárkou, dochází nutně k zaokrouhlování. Důsledkem toho je, že dostaneme pouze přibližné řešení \tilde{x} . Zásadní otázkou je, nakolik je tento výsledek přesný.

Celá tato problematika je technicky obtížná a výsledky formulované jako přesné matematické věty jsou složité, zájemci je naleznou ve specializované literatuře. My se omezíme na volný popis základních poznatků, které ukazují, co ovlivňuje přesnost numericky nalezeného řešení. V oddílu 1.4 jsme mluvili o podmíněnosti numerických úloh a v oddílu 1.5 pak o numericky stabilních algoritmech. Jak uvidíme, právě tyto vlastnosti hrají při numerickém řešení soustav lineárních rovnic klíčovou roli.

Numerická stabilita

Přibližné řešení \tilde{x} soustavy $Ax = B$ nebude tuto soustavu splňovat přesně. Budeme však požadovat, aby splňovalo přesně soustavu, jejíž koeficienty se jen málo liší od koeficientů

matice A a vektoru B . Proto zavádíme následující pojem.

Definice 3.19 Algoritmus pro výpočet řešení lineární soustavy $Ax = B$ se nazývá *numericky stabilní*, jestliže přibližné řešení \tilde{x} jím vypočtené splňuje soustavu

$$(A + \delta A)\tilde{x} = B + \delta B,$$

kde matice δA a vektor δB jsou malé (v nějakých normách). Přitom δA se nazývá *chybová matice*.

Je možné, že některý algoritmus obecně není numericky stabilní, ale pro jisté speciální případy numericky stabilní je. Lze ukázat (viz [22, str. 121], [52, str. 215] a v nich uvedené odkazy), že pro metody, o kterých jsme mluvili v předchozím textu, platí:

GEM bez výběru hlavních prvků	numericky nestabilní
GEM s částečným nebo úplným výběrem hlavních prvků	numericky stabilní
GEM bez výběru hlavních prvků pro ryze diagonálně dominantní nebo pozitivně definitní matice	numericky stabilní
Choleského metoda	numericky stabilní

Předchozí závěry platí i pro metodu LU resp. PLU rozkladu, protože je ekvivalentní GEM bez výběru hlavních prvků resp. s částečným výběrem hlavních prvků.

Je-li $\tilde{\mathbf{x}}$ přibližné řešení soustavy $A\mathbf{x} = B$, nazývá se rozdíl $\mathbf{r} = B - A\tilde{\mathbf{x}}$ *reziduum*. Předchozí poznatky lze stručně shrnout takto:

GEM s částečným nebo úplným výběrem hlavních prvků zaručuje vznik malých reziduí.

Otázkou ovšem je, zda malé reziduum zajišťuje, že $\tilde{\mathbf{x}}$ je dobrou aproximací přesného řešení \mathbf{x} . Odpověď je bohužel negativní, jak ukazují už příklady dvou rovnic o dvou neznámých.

Podmíněnost soustav lineárních rovnic

Je-li algoritmus pro řešení soustavy $A\mathbf{x} = B$, mající přesné řešení \mathbf{x} , numericky stabilní, vyhovuje získané přibližné řešení $\tilde{\mathbf{x}}$ soustavě, která má od původní soustavy jen velmi málo odlišné koeficienty (definice 3.19). Otázkou však je, zda naopak platí, že řešení soustavy s málo odlišnými koeficienty od soustavy $A\mathbf{x} = B$ se jen málo liší od jejího přesného řešení \mathbf{x} .

K téže otázce můžeme dojít i jinak. Koeficienty matice A a vektoru B jsou často získané měřením a jsou tudíž jen přibližné. Říkáme, že prvky A a B vznikly poruchou neboli perturbací přesných hodnot. Ve skutečnosti tudíž neřešíme „správnou“ soustavu, ale pouze její aproximaci. Musíme se tedy zajímat o to, jak vzniklé nepřesnosti ovlivní řešení a zda vůbec přesné řešení perturbovaného systému dobře aproximuje přesné řešení systému se správnými, ale nám neznámými koeficienty.

Uvažujme soustavu $A\mathbf{x} = B$, která má přesné řešení \mathbf{x} . Jejím porušením vznikne soustava $\tilde{A}\tilde{\mathbf{x}} = \tilde{B}$, mající přesné řešení $\tilde{\mathbf{x}}$. Označme $\tilde{A} = A + \delta A$, $\tilde{B} = B + \delta B$ a $\tilde{\mathbf{x}} = \mathbf{x} + \delta \mathbf{x}$. Tedy δA a δB jsou poruchy matice A a vektoru B , $\delta \mathbf{x}$ udává, o kolik se liší přesné řešení porušené soustavy $(A + \delta A)(\mathbf{x} + \delta \mathbf{x}) = B + \delta B$ od přesného řešení soustavy $A\mathbf{x} = B$.

Připomeňme, že jsme v definici 1.19 zavedli číslo podmíněnosti $\kappa(A)$ regulární matice A dané vztahem $\kappa(A) = \|A\| \|A^{-1}\|$. Vždy platí $\kappa(A) \geq 1$. O normách, které budeme v dalším potřebovat, budeme předpokládat, že splňují požadavky uvedené před zmíněnou definicí. Lze dokázat následující výsledek (viz [22, str. 125]).

Věta 3.20 *Nechť A je regulární matice a pro poruchu δA platí $\|\delta A\| < 1/\|A^{-1}\|$ (tj. δA je dostatečně malá porucha). Potom pro řešení \mathbf{x} soustavy $A\mathbf{x} = \mathbf{B}$ a řešení $\tilde{\mathbf{x}} = \mathbf{x} + \delta\mathbf{x}$ porušené soustavy $(A + \delta A)(\mathbf{x} + \delta\mathbf{x}) = \mathbf{B} + \delta\mathbf{B}$ platí*

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta\mathbf{B}\|}{\|\mathbf{B}\|} + \frac{\|\delta A\|}{\|A\|} \right). \quad (3.10)$$

Předchozí vztah udává horní odhad relativní chyby řešení \mathbf{x} . Odhad závisí na relativních chybách A a B a číslu podmíněnosti $\kappa(A)$.

Všimneme si dvou speciálních případů. Předpokládejme nejprve, že porušen je jen vektor B , tj. $\delta A = O$, takže $\|\delta A\| = 0$. Vyjde

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa(A) \frac{\|\delta\mathbf{B}\|}{\|\mathbf{B}\|}. \quad (3.11)$$

Předchozí vztah ukazuje, že relativní chyba řešení \mathbf{x} nemůže být víc než $\kappa(A)$ -krát větší, než je relativní chyba pravé strany B . Existují však matice A , pro něž v (3.11) nastává rovnost, tudíž relativní chyba řešení může být až $\kappa(A)$ -krát větší než je relativní chyba pravé strany. Tedy číslo podmíněnosti funguje jako jakýsi zesilující koeficient, který nás informuje, jak moc se může zvýšit relativní chyba řešení oproti relativní chybě pravé strany. Pokud je tedy $\kappa(A)$ hodně velké, může se stát, že i když je relativní chyba vektoru B malá, může být relativní chyba řešení \mathbf{x} hodně velká.

Pro reziduum platí

$$\mathbf{r} = B - A\tilde{\mathbf{x}} = B - (B + \delta B) = -\delta B,$$

tedy

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa(A) \frac{\|\mathbf{r}\|}{\|B\|}.$$

Tato nerovnost potvrzuje skutečnost, kterou jsme již konstatovali: Je-li reziduum malé, nemusí být (v případě velkého čísla podmíněnosti $\kappa(A)$) malá relativní chyba řešení.

Nyní předpokládejme, že je porušena jen matice soustavy A , tj. $\delta B = \mathbf{0}$, takže $\|\delta B\| = 0$. Tentokrát vyjde

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \frac{\|\delta A\|}{\|A\|}} \frac{\|\delta A\|}{\|A\|}. \quad (3.12)$$

Protože $0 < 1 - \kappa(A) \frac{\|\delta A\|}{\|A\|} < 1$, je

$$\frac{\kappa(A)}{1 - \kappa(A) \frac{\|\delta A\|}{\|A\|}} \frac{\|\delta A\|}{\|A\|} > \kappa(A) \frac{\|\delta A\|}{\|A\|}.$$

Obdobně jako v předchozí situaci i v tomto případě může být relativní chyba řešení mnohem větší, než je relativní chyba matice soustavy A .

Číslo podmíněnosti matice $\kappa(A)$ tedy hraje důležitou roli při numerickém řešení lineárních soustav $A\mathbf{x} = B$. Z předchozích úvah je vidět, že soustavy se špatně podmíněnou maticí (např. $\kappa(A) > 100$) dělají při numerickém řešení problémy a nezáleží přitom na volbě algoritmu. Jedinou cestou je použití mantisy s velkým počtem cifer; ani hardwarová dvojnásobná přesnost však nemusí být dostatečná. Naopak soustavy s dobře podmíněnou

maticí soustavy lze GEM s částečným nebo úplným výběrem hlavních prvků úspěšně řešit.

Poznámka 3.21 Poznamenejme, že řešení \mathbf{x} soustavy $A\mathbf{x} = B$ s regulární maticí A (mající tedy jediné řešení) spojitě závisí na A a B . To vyplývá ze vztahu $\mathbf{x} = A^{-1}B$, vzorce pro inverzní matici $A^{-1} = \frac{1}{\det A} \text{adj } A^T$ a definičního vzorce pro determinant; připomeňme, že prvky adjungované matice jsou minory matice A , tj. jsou rovněž definované pomocí determinantů. Úloha nalezení řešení soustavy $A\mathbf{x} = B$ je tudíž korektní (viz oddíl 1.4), avšak nemusí být dobře podmíněná.

3.2 Iterační metody

Jak již bylo řečeno, u těchto metod konstruujeme posloupnost vektorů $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$, vlastně jakýchsi přibližných řešení, která konverguje k přesnému řešení systému $A\mathbf{x} = B$. Celá situace připomíná postup při řešení nelineárních rovnic z kapitoly 2. Musíme provést následující kroky:

- Upravit výchozí soustavu $A\mathbf{x} = B$ na tvar, který umožní vytvářet ze zvolené počáteční aproximace $\mathbf{x}^{(0)}$ další členy $\mathbf{x}^{(k)}$.
- Ověřit konvergenci posloupnosti $\{\mathbf{x}^{(k)}\}$ k přesnému řešení.
- Po konečném počtu kroků na základě nějakého kritéria (zastavovací podmínky) rozhodnout o ukončení postupu. Poslední člen získané posloupnosti prohlásíme za přibližné řešení.

Iterační metody se používají zejména pro řešení rozsáhlých systémů s řídkými maticemi soustavy. Takové systémy se vyskytují v řadě aplikací. Pro menší systémy (počet rovnic je kolem stovek) obvykle dáváme přednost přímým metodám, které jsou většinou přesnější.

Úprava na tvar vhodný k iterování

Máme zadán lineární systém $A\mathbf{x} = B$ s regulární maticí soustavy $A \in \mathbb{M}_n$, vektorem pravých stran $B \in \mathbb{R}^n$ a vektorem neznámých $\mathbf{x} \in \mathbb{R}^n$. Matici A vyjádříme ve tvaru $M + N$, kde M je regulární, takže existuje M^{-1} . Po úpravách dostaneme:

$$\begin{aligned} A\mathbf{x} &= B, \\ (M + N)\mathbf{x} &= B, \\ M\mathbf{x} &= -N\mathbf{x} + B, \\ \mathbf{x} &= -M^{-1}N\mathbf{x} + M^{-1}B. \end{aligned}$$

Označíme-li $T = -M^{-1}N$ a $C = M^{-1}B$, platí

$$\mathbf{x} = T\mathbf{x} + C, \quad T \in \mathbb{M}_n, C \in \mathbb{R}^n.$$

Soustava $\mathbf{x} = T\mathbf{x} + C$ je ekvivalentní s původní soustavou $A\mathbf{x} = B$. Nyní zvolíme počáteční aproximaci $\mathbf{x}^{(0)} \in \mathbb{R}^n$ a definujeme $(k + 1)$ -ní aproximaci řešení vztahem

$$\mathbf{x}^{(k+1)} = T\mathbf{x}^{(k)} + C, \quad k = 0, 1, 2, \dots \quad (3.13)$$

Všimneme si podmínek zaručujících konvergenci. Ve formulaci následujícího výsledku budeme potřebovat pojem spektrálního poloměru matice zavedený v definici 1.20.

Věta 3.22 *Posloupnost určená vzorcem (3.13) konverguje pro libovolnou počáteční aproximaci $\mathbf{x}^{(0)}$ právě tehdy, když $\rho(T) < 1$.*

Důkaz viz [22, str. 135] nebo [52, str. 624]. Čím menší je $\rho(T)$, tím je konvergence rychlejší. Je-li $\rho(T)$ jen o málo menší než jedna, může se díky zaokrouhlovacím chybám stát, že dostaneme nekonzerventní posloupnost.

Určit spektrální poloměr matice je obtížné. Nicméně s využitím věty 1.21 dostaneme následující postačující podmínku konvergence, která se ověří daleko snáz.

Důsledek 3.23 *Je-li $\|T\| < 1$, kde $\|\cdot\|$ je libovolná maticová norma, která je souhlasná s nějakou vektorovou normou, konverguje iterační proces (3.13) pro libovolnou počáteční aproximaci $\mathbf{x}^{(0)}$.*

Zastavovací podmínky

Protože při numerickém řešení můžeme provést pouze konečný počet kroků, musíme výpočet v jistém okamžiku zastavit. To učiníme, když na základě vhodného kritéria považujeme získané řešení za dostatečně přesné. Popíšeme nyní, jak v praxi postupujeme.

Uvažujme soustavu $A\mathbf{x} = B$, jejíž řešení získáváme iterační metodou (3.13). Zvolíme nějaké malé číslo $\varepsilon > 0$ (*přesnost řešení*) a vhodnou vektorovou a s ní souhlasnou maticovou normu. V praxi se používají následující *zastavovací podmínky*.

$$\begin{aligned} \text{i)} \quad & \frac{\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|}{\|\mathbf{x}^{(k)}\|} < \varepsilon, \\ \text{ii)} \quad & \frac{\|\mathbf{r}^{(k)}\|}{\|A\|\|\mathbf{x}^{(k)}\| + \|B\|} < \varepsilon, \quad \text{kde } \mathbf{r}^{(k)} = B - A\mathbf{x}^{(k)}. \end{aligned}$$

V prvním případě tedy posuzujeme relativní změnu aproximace, v druhém relativní velikost rezidua. Je-li zvolená podmínka splněna, výpočet ukončíme a $\mathbf{x}^{(k)}$ považujeme za přibližné řešení.

Jsou-li splněny předpoklady důsledku 3.23, lze dokázat nerovnost (viz [22, str. 136])

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|T\|^k}{1 - \|T\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|.$$

Tentokrát jsme tedy schopni odhadnout vzdálenost k -té aproximace od přesného řešení a to pomocí vzdálenosti prvních dvou aproximací a normy matice T . Protože $\lim_{k \rightarrow \infty} \|T\|^k = 0$, je pro dostatečně velké k pravá strana předchozí nerovnosti libovolně malá, takže lze najít takové k , že bude menší než zvolené ε . Tento odhad je však obvykle hodně pesimistický, skutečná vzdálenost $\|\mathbf{x} - \mathbf{x}^{(k)}\|$ bývá pro takto určené k mnohem menší, tedy k je zbytečně velké.

Nyní si všimneme konkrétních úprav systému $A\mathbf{x} = B$ na tvar vhodný k iteraci.

3.2.1 Jacobiova metoda

Uvažujme systém $Ax = B$ s regulární maticí soustavy A . Tuto matici rozdělíme na tři části: diagonálu a dvě trojúhelníkové matice s naddiagonálními resp. poddiagonálními prvky (s opačnými znaménky).

Označíme

$$D = \begin{pmatrix} a_{11} & 0 & \dots & 0 & 0 \\ 0 & a_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & a_{n-1,n-1} & 0 \\ 0 & 0 & \dots & 0 & a_{nn} \end{pmatrix}, \quad (3.14)$$

$$L = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ -a_{21} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_{n-1,1} & -a_{n-1,2} & \dots & 0 & 0 \\ -a_{n1} & -a_{n2} & \dots & -a_{n,n-1} & 0 \end{pmatrix}, \quad (3.15)$$

$$U = \begin{pmatrix} 0 & -a_{12} & \dots & -a_{1,n-1} & -a_{1n} \\ 0 & 0 & \dots & -a_{2,n-1} & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -a_{n-1,n} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}. \quad (3.16)$$

Zřejmě platí $A = D - L - U$. Budeme předpokládat, že matice A má nenulové diagonální prvky. Potom je matice D regulární, k ní inverzní matice D^{-1} je rovněž diagonální a má na diagonále prvky $1/a_{11}, \dots, 1/a_{nn}$.

Tvar vhodný k iteraci získáme tak, že z první rovnice osamostatníme neznámou x_1 ,

z druhé rovnice neznámou x_2 , a tak dále, až z poslední rovnice osamostatníme neznámou x_n . Maticově vypadá zápis takto:

$$\begin{aligned} Ax &= B, \\ (D - L - U)x &= B, \\ Dx &= (L + U)x + B, \\ x &= D^{-1}(L + U)x + D^{-1}B, \\ x &= T_J x + D^{-1}B, \end{aligned}$$

kde

$$T_J = D^{-1}(L + U) = \begin{pmatrix} 0 & -\frac{a_{12}}{a_{11}} & \cdots & -\frac{a_{1,n-1}}{a_{11}} & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \cdots & -\frac{a_{2,n-1}}{a_{22}} & -\frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \cdots & -\frac{a_{n,n-1}}{a_{nn}} & 0 \end{pmatrix}.$$

Iterační formule bude mít tvar

$$\mathbf{x}^{(k+1)} = T_J \mathbf{x}^{(k)} + D^{-1}B, \quad k = 0, 1, 2, \dots \quad (3.17)$$

Konvergence Jacobiovy metody

Podle věty 3.22 platí následující tvrzení.

Věta 3.24 *Jacobiova metoda (3.17) konverguje pro libovolnou počáteční aproximaci k řešení soustavy $A\mathbf{x} = B$ právě tehdy, když $\rho(T_J) = \rho(D^{-1}(L + U)) < 1$.*

Určení spektrálního poloměru iterační matice je však obtížné. Ve speciálních případech existuje jednodušší postačující podmínka konvergence (viz [8, str. 213], [22, str. 140] nebo [52, str. 625]).

Věta 3.25 *Nechť je splněna některá z následujících podmínek:*

- 1) *Matice A je ryze řádkově diagonálně dominantní.*
- 2) *Matice A je ryze sloupcově diagonálně dominantní.*

Pak Jacobiova metoda (3.17) konverguje pro libovolnou počáteční aproximaci k řešení soustavy $A\mathbf{x} = B$.

3.2.2 Gaussova-Seidelova metoda

U Jacobiovy metody je zřejmé, že pokud známe všechny složky k -té iterace $\mathbf{x}^{(k)}$, můžeme počítat složky následující iterace $\mathbf{x}^{(k+1)}$ v libovolném pořadí, nemusí to být nejprve $x_1^{(k+1)}$, pak $x_2^{(k+1)}$ atd. Pokud má počítač více procesorů, lze toho případně využít k paralelním výpočtům jednotlivých složek, každá se počítá na jiném procesoru, čímž se výpočet může významně urychlit.

Na druhé straně, počítáme-li složky iterace $\mathbf{x}^{(k+1)}$ postupně v pořadí $x_1^{(k+1)}$, pak $x_2^{(k+1)}$ atd., je přirozeným nápadem využít při výpočtu jednotlivých složek ty složky $(k+1)$ -ní iterace, které jsou již spočtené. Očekáváme totiž, že dávají přesnější aproximaci než tytéž složky k -té iterace. Tedy při výpočtu $x_2^{(k+1)}$ použijeme ve vzorci (3.18) aproximaci $x_1^{(k+1)}$ místo $x_1^{(k)}$, při výpočtu $x_3^{(k+1)}$ použijeme $x_1^{(k+1)}$ a $x_2^{(k+1)}$ místo $x_1^{(k)}$ a $x_2^{(k)}$ atd.

V maticovém zápisu vypadá uvedený postup takto:

$$\begin{aligned}Ax &= B, \\(D - L - U)x &= B, \\(D - L)x &= Ux + B, \\x &= (D - L)^{-1}Ux + (D - L)^{-1}B, \\x &= T_{GS}x + (D - L)^{-1}B,\end{aligned}$$

kde

$$T_{GS} = (D - L)^{-1}U.$$

Všimněte si, že matice $D - L$ je regulární, protože je dolní trojúhelníková a na diagonále má nenulové prvky, tudíž její inverzní matice existuje a je rovněž dolní trojúhelníková.

Iterační formule bude mít tvar

$$\mathbf{x}^{(k+1)} = T_{GS}\mathbf{x}^{(k)} + (D - L)^{-1}B, \quad k = 0, 1, 2, \dots \quad (3.19)$$

Tato úprava se nazývá *Gaussova¹-Seidelova² iterační metoda*. K iterování však používáme tvar

$$\mathbf{x}^{(k+1)} = D^{-1}L\mathbf{x}^{(k+1)} + D^{-1}U\mathbf{x}^{(k)} + D^{-1}B, \quad k = 0, 1, 2, \dots \quad (3.20)$$

Zřejmě i -tá složka má tvar

$$x_i^{(k+1)} = - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad i = 1, 2, \dots, n. \quad (3.21)$$

Tato podoba je názornější a je z ní lépe vidět, že odpovídá předchozímu slovnímu popisu. Rozepsaná soustava po úpravě na tvar vhodný k iterování je potom

$$\begin{aligned} x_1^{(k+1)} &= && - \frac{a_{12}}{a_{11}} x_2^{(k)} && - \dots - \frac{a_{1,n-1}}{a_{11}} x_{n-1}^{(k)} && - \frac{a_{1n}}{a_{11}} x_n^{(k)} && + \frac{b_1}{a_{11}}, \\ x_2^{(k+1)} &= - \frac{a_{21}}{a_{22}} x_1^{(k+1)} && && - \dots - \frac{a_{2,n-1}}{a_{22}} x_{n-1}^{(k)} && - \frac{a_{2n}}{a_{22}} x_n^{(k)} && + \frac{b_2}{a_{22}}, \\ &\dots && && && && \\ x_n^{(k+1)} &= - \frac{a_{n1}}{a_{nn}} x_1^{(k+1)} && - \frac{a_{n2}}{a_{nn}} x_2^{(k+1)} && - \dots - \frac{a_{n,n-1}}{a_{nn}} x_{n-1}^{(k+1)} && && + \frac{b_n}{a_{nn}}. \end{aligned}$$

¹Viz str. 207.

²**Philipp Ludwig von Seidel** (1821–1896) (čti zajdl) — německý astronom a matematik.

Konvergence Gaussovy-Seidelovy metody

Podle věty 3.22 platí následující tvrzení.

Věta 3.26 *Gaussova-Seidelova metoda (3.19) konverguje pro libovolnou počáteční aproximaci k řešení soustavy $Ax = B$, právě když $\rho(T_{GS}) = \rho((D - L)^{-1}U) < 1$.*

Jak jsme již konstatovali, určení spektrálního poloměru iterační matice je obtížné. I pro tuto metodu ve speciálních případech existují jednodušší postačující podmínky konvergence.

Věta 3.27 *Nechť je splněna některá z následujících podmínek:*

- 1) *Matice A je ryze řádkově diagonálně dominantní (viz [8, str. 213], [22, str. 142] nebo [52, str. 628]).*
- 2) *Matice A je ryze sloupcově diagonálně dominantní (viz [8, str. 213] nebo [14, str. 351]).*
- 3) *Matice A je symetrická pozitivně definitní (viz [8, str. 211], [14, str. 427] nebo [52, str. 631]).*

Pak Gaussova-Seidelova metoda (3.19) konverguje pro libovolnou počáteční aproximaci k řešení soustavy $Ax = B$.

Srovnání Jacobiovy a Gaussovy-Seidelovy metody

Množiny soustav $Ax = B$, pro něž konverguje Jacobiova resp. Gaussova-Seidelova metoda, jsou nesrovnatelné. Existují soustavy, pro něž Jacobiova metoda konverguje a Gaussova-Seidelova metoda nekonverguje, a rovněž soustavy, pro něž Gaussova-Seidelova metoda konverguje a Jacobiova metoda nekonverguje. Rovněž nemusí být pravda, že pokud obě tyto metody konvergují, je Gaussova-Seidelova metoda rychlejší, tj. že platí $\rho(T_{GS}) < \rho(T_J) < 1$ (menší spektrální poloměr znamená rychlejší konvergenci).

Nicméně je řada důležitých a často se vyskytujících případů, kdy tomu tak je. Například platí, že pokud má matice A kladné diagonální prvky a nekladné mimodiagonální prvky a Jacobiova metoda konverguje, konverguje i Gaussova-Seidelova metoda a $0 < \rho(T_{GS}) < \rho(T_J) < 1$ nebo $\rho(T_{GS}) = \rho(T_J) = 0$ (viz [52, str. 629], [55, str. 76]).

Z vět 3.25 a 3.27 víme, že pokud je matice soustavy A ryze řádkově nebo sloupcově diagonálně dominantní, konverguje Jacobiova i Gaussova-Seidelova metoda. Ukazuje se, že tento požadavek lze oslabit, stačí když v definici 3.2 budou nerovnosti splněny jako neostré s výjimkou aspoň jedné, která bude ostrá. Jde tedy o jakousi neryzí verzi diagonální dominantnosti.

Matice soustavy však musí mít navíc jistou speciální vlastnost. Připomeňme, že permutační matice je matice, která vznikne z jednotkové matice E přeházením některých řádků. Násobíme-li

nějakou matici zleva (zprava) permutační maticí, vymění se v této matici odpovídajícím způsobem její řádky (sloupce).

Definice 3.28 Řekneme, že čtvercová matice $A \in \mathbb{M}_n$ je *ireducibilní*, jestliže neexistuje permutační matice P taková, že platí

$$P^T A P = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ O & \tilde{A}_{22} \end{pmatrix},$$

kde \tilde{A}_{11} je matice velikosti $p \times p$ a \tilde{A}_{22} je matice velikosti $q \times q$, přičemž $p + q = n$, $p \geq 1$, $q \geq 1$.

Jinými slovy, nesmí být možné vytvořit z matice A přeházením některých řádků a týchž sloupců matici, která má v levém dolním rohu blok nul o velikosti $q \times p$, kde $p + q = n$. Podrobněji viz např. [18, str. 71], kde je popsáno, jak lze tuto vlastnost ověřit.

Věta 3.29 *Uvažujme soustavu $Ax = B$. Nechť matice A je ireducibilní a platí*

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, \dots, n,$$

přičemž aspoň pro jeden index i_0 platí dokonce

$$|a_{i_0 i_0}| > \sum_{\substack{j=1 \\ j \neq i_0}}^n |a_{i_0 j}|.$$

Pak Jacobiova metoda (3.17) konverguje pro libovolnou počáteční aproximaci k řešení soustavy $Ax = B$.

Předchozí tvrzení platí i pro sloupcové diagonální dominování a rovněž pro Gaussovu-Seidelovu metodu. Viz [8, str. 213], [52, str. 626 a 628] a [55, str. 79].

Soustavy splňující předpoklady předchozí věty se často vyskytují v souvislosti s numerickým řešením parciálních diferenciálních rovnic.

3.2.3 Relaxační metody, metoda SOR

Relaxační metody jsou založeny na myšlence vyjít z některé iterační metody a upravit iterační matici tak, aby nová metoda konvergovala rychleji (nebo vůbec konvergovala, pokud výchozí metoda nekonverguje). Princip si vysvětlíme na Gaussově-Seidelově metodě, což je nejčastější případ, lze ho však použít i pro Jacobiovu metodu (viz [13, str. 34]).

Postup relaxace lze popsat takto:

- 1) Složky $(k + 1)$ -ní iterace počítáme postupně, tj. $x_1^{(k+1)}$, $x_2^{(k+1)}$ atd.
- 2) Při výpočtu složky $x_i^{(k+1)}$ nejprve předpovíme její hodnotu základní metodou, čímž získáme jistou hodnotu $\tilde{x}_i^{(k+1)}$.
- 3) Složku $x_i^{(k+1)}$ vypočteme z hodnot $\tilde{x}_i^{(k+1)}$ a $x_i^{(k)}$ jako jejich vážený průměr. Tedy

$$x_i^{(k+1)} = \omega \tilde{x}_i^{(k+1)} + (1 - \omega)x_i^{(k)}, \quad i = 1, 2, \dots, n.$$

Číslo ω se nazývá *relaxační parametr*. Je-li $\omega < 1$, mluvíme o *dolní relaxaci*, je-li $\omega > 1$, mluvíme o *horní relaxaci*. Pro $\omega = 1$ dostaneme základní metodu. Dolní relaxace se často používá pro zajištění konvergence nekonvergentních metod, horní relaxace pak pro

zrychlení konvergence. Relaxace Gaussovy-Seidelovy metody se nazývá *metoda SOR* (Successive Overrelaxation). Navrhli ji téměř současně v roce 1950 [David M. Young, Jr.](#) a [Stanley Phillips Frankel](#). Viz [64].

V případě relaxace Gaussovy-Seidelovy metody, tj. metody SOR, dostaneme podle vzorce (3.21), že

$$\tilde{x}_i^{(k+1)} = - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad i = 1, 2, \dots, n.$$

Na pravé straně předchozího výrazu vystupují hodnoty $x_j^{(k+1)}$, protože pro $j < i$ jsou již známé. Tedy

$$\begin{aligned} x_i^{(k+1)} &= \omega \tilde{x}_i^{(k+1)} + (1 - \omega)x_i^{(k)} = \\ &= \omega \left(- \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}} \right) + (1 - \omega)x_i^{(k)} = \\ &= x_i^{(k)} + \omega \left(- \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - x_i^{(k)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}} \right). \end{aligned} \quad (3.22)$$

Odvodíme rovněž maticový zápis. Ze vztahu (3.20) po vynásobení maticí D dostaneme pro předpovězenou hodnotu

$$D\tilde{\mathbf{x}}^{(k+1)} = L\mathbf{x}^{(k+1)} + U\mathbf{x}^{(k)} + B, \quad k = 0, 1, 2, \dots$$

Tedy

$$\begin{aligned} D\mathbf{x}^{(k+1)} &= \omega D\tilde{\mathbf{x}}^{(k+1)} + (1 - \omega)D\mathbf{x}^{(k)} = \\ &= \omega(L\mathbf{x}^{(k+1)} + U\mathbf{x}^{(k)} + B) + (1 - \omega)D\mathbf{x}^{(k)} = \\ &= \omega L\mathbf{x}^{(k+1)} + [(1 - \omega)D + \omega U]\mathbf{x}^{(k)} + \omega B, \end{aligned}$$

takže

$$\begin{aligned} (D - \omega L)\mathbf{x}^{(k+1)} &= [(1 - \omega)D + \omega U]\mathbf{x}^{(k)} + \omega B, \\ \mathbf{x}^{(k+1)} &= (D - \omega L)^{-1}[(1 - \omega)D + \omega U]\mathbf{x}^{(k)} + \omega(D - \omega L)^{-1}B, \\ \mathbf{x}^{(k+1)} &= T_\omega\mathbf{x}^{(k)} + \omega(D - \omega L)^{-1}B, \end{aligned} \tag{3.23}$$

kde

$$T_\omega = (D - \omega L)^{-1}[(1 - \omega)D + \omega U].$$

Matice $D - \omega L$ je dolní trojúhelníková a má na diagonále nenulové prvky a_{11}, \dots, a_{nn} , takže je regulární, a tudíž existuje její inverze.

Zajímá nás, pro která ω je metoda SOR daná vztahem (3.23) konvergentní a kdy bude tato konvergence nejrychlejší, tj. kdy bude spektrální poloměr iterační matice T_ω nejmenší.

Ukazuje se však, že má smysl uvažovat pouze některé hodnoty ω . Platí totiž následující tvrzení (viz [52, str. 631]).

Věta 3.30 (Kahan) *Pro libovolnou matici A platí $\rho(T_\omega) \geq |\omega - 1|$.*

Podle věty 3.22 víme, že metoda SOR bude konvergovat, právě když $\rho(T_\omega) < 1$. Tudíž aby metoda SOR mohla konvergovat, musí platit $|\omega - 1| < 1$, tj. $\omega \in (0, 2)$. Tato podmínka je pouze nutná (pokud není splněna, metoda SOR nekonverguje), nemusí však zaručovat konvergenci. Jak ukazuje následující tvrzení, je však postačující v případě pozitivně definitních matic (viz [8, str. 211], [52, str. 631] nebo [55, str. 83]).

Věta 3.31 (Ostrowski, Reich) *Je-li A symetrická pozitivně definitní matice, platí $\rho(T_\omega) < 1$ pro libovolné $\omega \in (0, 2)$, tj. metoda SOR konverguje.*

Určit v obecném případě, pro které $\omega \in (0, 2)$ je metoda SOR konvergentní, není jednoduché a ještě obtížnější je určit, pro které ω je $\rho(T_\omega)$ nejmenší, tj. konvergence je nejrychlejší. Pro některé speciální typy matic jsou hodnoty známé, viz např. [52, str. 633] a [55, str. 111]. Existují postupy, jak relaxační parametr ω průběžně měnit, zmenšovat $\rho(T_\omega)$ a tím urychlovat konvergenci.

Poznámka 3.32 Uvedené iterační metody jsou tzv. klasické. Dnes jsou známé jiné, účinnější metody. Jejich výklad však přesahuje možnosti tohoto úvodního textu. Patří k nim metody založené na minimalizaci kvadratického funkcionálu, např. *metoda největšího spádu*, viz [22, str. 114]. Významné jsou metody založené na *Krylovových¹ prostorech*. K nim patří *metoda sdružených gradientů*, kterou lze použít pro symetrické pozitivně definitní matice, pro libovolné matice pak *metoda bikonjugovaných gradientů*, metoda GMRES (generalized minimum residual method), metoda QMR (quasi-minimal residual method) a další. Viz např. [52, str. 657]. Jde vlastně o přímé metody (v přesné aritmetice), ale v důsledku zaokrouhlovacích chyb dochází ke vzniku nekonečné posloupnosti aproxi-

¹ **Alexej Nikolajevič Krylov** (1863–1945) — ruský lodní inženýr a matematik. Zabýval se aplikovanou matematikou.

mací. Výpočet musíme ukončit pomocí vhodné zastavovací podmínky. Tyto metody se používají zejména na rozsáhlé systémy s řídkými maticemi.

Pojmy k zapamatování

- soustava lineárních rovnic
- existence a jednoznačnost řešení
- přímé metody
- iterační metody
- Gaussova eliminační metoda, přímý a zpětný chod
- částečný a úplný výběr hlavních prvků v GEM
- výpočtová náročnost GEM
- řádkově a sloupcově diagonálně dominantní matice
- symetrické pozitivně definitní matice
- metoda LU rozkladu, částečný výběr hlavních prvků

- Choleského metoda
- numericky stabilní algoritmy
- podmíněnost soustav lineárních rovnic
- tvar systému lineárních rovnic vhodný k iterování
- podmínky konvergence iteračních metod
- zastavovací podmínky
- Jacobiova metoda
- Gaussova-Seidelova metoda
- relaxační metody, metoda SOR

Kontrolní otázky

1. Kdy má čtvercová soustava rovnic $Ax = B$ právě jedno řešení?
2. Uveďte ekvivalentní podmínky, které zaručují, že matice A je regulární.
3. Co lze říci o matici A , víte-li, že k ní existuje matice inverzní A^{-1} ?
4. Kolik řešení má systém $Ax = B$ v případě, že matice A je regulární?
5. Které úpravy patří mezi elementární řádkové úpravy matic?
6. Proč je výhodné upravit soustavu $Ax = B$ na soustavu s ní ekvivalentní $Cx = D$, kde C je horní trojúhelníková matice?
7. Jaké je základní dělení numerických metod pro řešení soustav lineárních rovnic?
8. Které metody pro řešení soustav lineárních rovnic řadíme mezi přímé metody?
9. Kdy se zpravidla pro řešení soustav používají přímé metody?
10. Která metoda je základem většiny přímých metod pro řešení soustav lineárních rovnic?
11. Jaké části má Gaussova eliminační metoda?
12. Popište detailně přímý chod Gaussovy eliminační metody.
13. Které prvky označujeme jako hlavní prvky neboli pivoty?

14. Jaká je výpočtová náročnost jednotlivých částí Gaussovy eliminační metody?
15. Kdy lze řešit soustavu $Ax = B$ pomocí algoritmu GEM bez výběru hlavních prvků?
16. Co se stane, jestliže při řešení soustavy $Ax = B$ pomocí algoritmu GEM bez výběru hlavních prvků bude některý z prvků $a_{kk}^{(k)} = 0$, kde $k = 1, \dots, n - 1$?
17. Kdy se používá algoritmus GEM s částečným výběrem hlavních prvků?
18. Jak funguje algoritmus GEM s částečným výběrem hlavních prvků?
19. Co je to permutační matice? K čemu slouží?
20. Kdy se používá algoritmus GEM s úplným výběrem hlavních prvků?
21. Jak funguje algoritmus GEM s úplným výběrem hlavních prvků?
22. Co je to LU rozklad regulární matice?
23. Je-li A regulární matice, existuje vždy její LU rozklad?
24. Uveďte podmínky, za kterých existuje LU rozklad regulární matice A .
25. Vysvětlete, jak najdeme LU rozklad dané matice.
26. Popište kroky metody LU rozkladu pro řešení soustav lineárních rovnic.
27. Kdy je výhodnější použít pro řešení soustav lineárních rovnic metodu LU rozkladu místo Gaussovy eliminační metody?

28. Kdy se používá LU rozklad s částečným výběrem hlavních prvků?
29. Popište nalezení LU rozkladu s částečným výběrem hlavních prvků.
30. Kdy lze systém rovnic $Ax = B$ řešit Choleského metodou?
31. Jak vypadá Choleského rozklad matice A ?
32. Jaký je princip řešení systému rovnic Choleského metodou?
33. Je Choleského rozklad speciálním případem LU rozkladu matice A ?
34. Dostaneme při použití přímých metod vždy přesný výsledek? Zdůvodněte odpověď.
35. Které matice nazýváme ryze řádkově (sloupcově) diagonálně dominantní?
36. Může být ryze řádkově (sloupcově) diagonálně dominantní matice singulární?
37. Které matice nazýváme symetrické?
38. Kdy se symetrická matice nazývá pozitivně definitní?
39. K čemu slouží Sylvestrovo kritérium?
40. Které matice se označují jako řídké?
41. Které matice se označují jako pásové?
42. Co rozumíme numerickou stabilitou algoritmu pro řešení soustav lineárních rovnic?

43. Co je to reziduum?
44. Které přímé metody pro numerické řešení soustav lineárních rovnic jsou numericky stabilní?
45. Které přímé metody pro numerické řešení soustav lineárních rovnic jsou numericky nestabilní?
46. Jaký vliv má číslo podmíněnosti matice soustavy na přesnost řešení?
47. Dostaneme numericky stabilním algoritmem vždy dobrou aproximaci přesného řešení?
48. Co je to dobře podmíněná soustava lineárních rovnic?
49. Kdy se zpravidla používají při řešení lineárních soustav rovnic iterační metody?
50. Které metody pro řešení soustav lineárních rovnic řadíme mezi iterační metody?
51. Jak postupujeme, když hledáme řešení soustavy pomocí iteračních metod?
52. Jak vypadá základní formule pro iterační metody řešení soustav lineárních rovnic?
53. Co je potřeba ověřit před použitím iteračních metod pro řešení soustav lineárních rovnic?
54. Co je nutnou a postačující podmínkou konvergence iteračních metod pro řešení soustav lineárních rovnic?
55. Co je postačující podmínkou konvergence iteračních metod pro řešení soustav lineárních rovnic?

56. S čím souvisí rychlost konvergence iteračních metod? Kdy je konvergence iteračních metod rychlejší?
57. Uveďte zastavující podmínky, které se používají pro ukončení iteračních metod pro řešení soustav lineárních rovnic.
58. Popište Jacobiovu metodu.
59. Uveďte postačující podmínky konvergence Jacobiovy metody.
60. Popište Gaussovu-Seidelovu metodu.
61. Uveďte postačující podmínky konvergence Gaussovy-Seidelovy metody.
62. Jaký je vztah mezi konverencí Jacobiovy a Gaussovy-Seidelovy metody?
63. Vysvětlete princip relaxačních metod.
64. Co je to metoda SOR?
65. Pro jaké hodnoty relaxačního parametru může metoda SOR konvergovat?
66. Uveďte postačující podmínku konvergence metody SOR.

Testy ke kapitole 3

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Je-li matice A regulární, pak

$$|A| = 0.$$

$$|A| = 1.$$

$$A = A^T.$$

$$|A| \neq 0.$$

2. (1b.) Numerická metoda řešení soustavy $Ax = B$, která vede po konečném počtu kroků k přesnému řešení (předpokládáme, že nedochází k zaokrouhlování), se nazývá
přímá metoda. iterační metoda.

Newtonova metoda.

metoda postupných aproximací.

3. (1b.) Jsou-li všechny horní rohové hlavní minory čtvercové matice A v soustavě $A\mathbf{x} = B$ nenulové, pak
- soustavu nelze řešit GEM bez výběru hlavních prvků.
 - lze soustavu řešit GEM bez výběru hlavních prvků.
 - algoritmus GEM bez výběru hlavních prvků zhavaruje.
 - je potřeba použít GEM s částečným nebo úplným výběrem hlavních prvků.
4. (1b.) Porovnejte výpočetní náročnost přímého a zpětného chodu u GEM.
- | | |
|---------------------------------|----------------------------|
| Oba jsou zhruba stejně náročné. | Obecně je nelze srovnat. |
| Přímý chod je náročnější. | Zpětný chod je náročnější. |
5. (1b.) Jestliže je absolutní hodnota libovolného diagonálního prvku větší než součet absolutních hodnot zbývajících prvků v témže řádku, pak se čtvercová matice nazývá
- | | |
|-------------------------------------|---------------------------------------|
| ryze řádkově diagonálně dominantní. | ryze sloupcově diagonálně dominantní. |
| pozitivně definitní. | symetrická. |

6. (1b.) Řešíme-li numericky soustavu lineárních rovnic, která má jedno řešení, pak počet rovnic musí být stejný jako počet neznámých. Navíc mohou být některé rovnice lineárně závislé.

počet lineárně nezávislých rovnic musí být menší než počet neznámých.

je-li počet rovnic větší, než je počet neznámých, lze vynechat libovolné rovnice tak, aby zbylý počet rovnic odpovídal počtu neznámých.

počet rovnic může být větší jak počet neznámých. Je-li větší, jsou rovnice lineárně závislé, takže lze některé vynechat a zbylé rovnice budou lineárně nezávislé a bude jich tolik, co neznámých.

7. (1b.) Vyberte, co platí pro symetrické matice.

$$|A| = 0$$

$$A^T = A.$$

$$A^{-1} = A.$$

Všechny jejich diagonální prvky jsou nenulové.

8. (1b.) Existuje nějaký vztah mezi GEM bez výběru hlavního prvku a metodou LU rozkladu?

Ne. Obě metody sice patří mezi přímé metody, ale nic dalšího je nespojuje.

Ne. Metodu LU rozkladu používáme pro soustavy s malým počtem rovnic, zatímco GEM se používá pro soustavy s velkým počtem rovnic.

Ano. Pokud neexistuje LU rozklad matice soustavy, tak zhavaruje GEM bez výběru hlavního prvku.

Ano. Pokud lze soustavu vyřešit metodou LU rozkladu, pak ji nelze vyřešit GEM.

9. (1b.) Kdy dáváme přednost iteračním metodám před přímými metodami?

V případech, kdy není matice soustavy symetrická.

V případech, kdy řešíme soustavy s velkým počtem rovnic, jejichž matice soustavy jsou řídké.

V případech, kdy neočekáváme přesný výsledek, ale stačí nám jen přibližný odhad.

V případech, kdy soustavy nejsou příliš rozsáhlé (obsahují jen několik stovek rovnic).

10. (1b.) Na čem jsou založeny relaxační metody?

Hodnotu $x_i^{(k+1)}$, tj. i -tou složku $(k + 1)$ -ní iterace, počítáme jako vážený průměr hodnot $x_i^{(k)}$ a $x_i^{(k-1)}$.

Hodnotu $x_i^{(k+1)}$, tj. i -tou složku $(k + 1)$ -ní iterace, počítáme jako vážený průměr hodnot $x_i^{(k)}$ a $x_i^{(k+1)}$.

Při výpočtu i -té složky $(k + 1)$ -ní iterace nejprve základní metodou předpovíme hodnotu $\tilde{x}_i^{(k+1)}$ a z ní a hodnoty $x_i^{(k-1)}$ určíme $x_i^{(k+1)}$ jako jejich vážený průměr.

Při výpočtu i -té složky $(k + 1)$ -ní iterace nejprve základní metodou předpovíme hodnotu $\tilde{x}_i^{(k+1)}$ a z ní a hodnoty $x_i^{(k)}$ určíme $x_i^{(k+1)}$ jako jejich vážený průměr.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (1b.) Je-li matice A regulární, pak

je její determinant nulový.

k ní existuje inverzní matice A^{-1} .

je její hodnost $h(A) < n$, kde n je počet řádků matice A .

k ní existuje transponovaná matice, přičemž platí $A^T = A$.

2. (1b.) Co je výsledkem přímého chodu GEM?

Převod systému $A\mathbf{x} = B$ na ekvivalentní systém $C\mathbf{x} = D$, kde C je dolní trojúhelníková matice, která má na diagonále jedničky.

Převod systému $A\mathbf{x} = B$ na ekvivalentní systém $C\mathbf{x} = D$, kde C je horní trojúhelníková matice.

Rozklad matice A systému $A\mathbf{x} = B$ na dolní a horní trojúhelníkovou matici.

Posloupnost vektorů $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$, která konverguje k přesnému řešení.

3. (1b.) Soustava je pomocí GEM upravena na trojúhelníkový tvar

$$\begin{aligned}x_1 + 2x_2 - x_3 &= 19, \\x_2 - 3x_3 &= 5, \\2x_3 &= 2.\end{aligned}$$

Její řešení je

$$\mathbf{x}^T = (3, 8, 1).$$

$$\mathbf{x}^T = (4, 8, 1).$$

$$\mathbf{x}^T = (16, 2, 1).$$

$$\mathbf{x}^T = (21, 2, 1).$$

4. (1b.) Je-li čtvercová matice A soustavy $A\mathbf{x} = B$ ryze řádkově nebo sloupcově diagonálně dominantní, pak lze tuto soustavu řešit GEM bez výběru hlavních prvků.

Ano.

Ne.

5. (1b.) Která z následujících přímých metod není numericky stabilní?

Metoda PLU rozkladu.

Choleského metoda.

GEM bez výběru hlavních prvků.

GEM s částečným nebo úplným výběrem hlavních prvků.

6. (1b.) Vyberte pravdivé tvrzení. Choleského metoda

patří mezi přímé metody řešení soustav lineárních rovnic. Jde o speciální případ GEM.

patří mezi iterační metody řešení soustav lineárních rovnic. Jde o speciální případ relaxační metody.

patří mezi přímé metody řešení soustav lineárních rovnic. Jde o speciální případ metody LU rozkladu.

patří mezi přímé metody řešení soustav lineárních rovnic. Metodu používáme v případě, že matice soustavy má všechny prvky kladné.

7. (1b.) Vyberte postup při řešení soustav lineárních rovnic tvaru $A\mathbf{x} = B$ iteračními metodami.

Soustavu přepíšeme na tvar vhodný k iteraci, zvolíme počáteční aproximaci řešení $\mathbf{x}^{(0)}$ a po určitém počtu kroků ukončíme postup a poslední aproximaci prohlásíme za přibližné řešení.

Soustavu přepíšeme na tvar vhodný k iteraci, zvolíme počáteční aproximaci $\mathbf{x}^{(0)}$, ověříme podmínku konvergence posloupnosti $\{\mathbf{x}^{(k)}\}$ k přesnému řešení, po určitém počtu kroků (na základě nějaké zastavovací podmínky) ukončíme postup a poslední aproximaci prohlásíme za přibližné řešení.

Určíme spektrální poloměr matice soustavy, abychom věděli, zda bude iterační proces konvergovat, stanovíme počáteční aproximaci řešení $\mathbf{x}^{(0)}$, zkonstruujeme konvergentní posloupnost $\{\mathbf{x}^{(k)}\}$, po určitém počtu kroků (na základě nějaké zastavovací podmínky) ukončíme postup a poslední aproximaci prohlásíme za přibližné řešení.

8. (1b.) Vyberte nutnou a postačující podmínku, aby posloupnost aproximací $\mathbf{x}^{(k+1)} = T\mathbf{x}^{(k)} + C$, kde $k = 0, 1, \dots$, konvergovala pro libovolnou počáteční aproximaci $\mathbf{x}^{(0)}$.

$$\rho(T) = 0.$$

$$\rho(T) = 1.$$

$$\rho(T) < 1.$$

$$\rho(T) \leq \varepsilon, \text{ kde } \varepsilon < 0,5 \text{ je kladné číslo.}$$

9. (1b.) Metoda SOR vznikne relaxací

Choleského metody.

Gaussovy-Seidelovy metody.

Jacobiovy metody.

Gaussovy eliminační metody.

10. (1b.) Která z následujících vlastností nezaručuje při řešení soustavy $Ax = B$ Jacobi-ovou metodou konvergenci? A je ryze řádkově diagonálně dominantní. A je ryze sloupcově diagonálně dominantní. A je symetrická pozitivně definitní.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (1b.) Je-li $A \in \mathbb{M}_n$, kde $n > 1$, regulární matice, pak má soustava $Ax = B$
nekonečně mnoho řešení. právě jedno řešení.
dvě řešení, z nichž jedno je triviální. n řešení, kde n je počet neznámých.
2. (1b.) Je-li soustava $Cx = D$ ekvivalentní s výchozí soustavou $Ax = B$, pak to
znamená, že
obě soustavy mají stejná řešení. platí $A = C$ a $B = D$.
3. (1b.) Je-li soustava pomocí GEM upravena na tvar

$$\begin{aligned}x_1 - 3x_2 + 4x_3 &= 5, \\6x_2 - x_3 &= 1, \\12x_3 &= 24,\end{aligned}$$

pak postup, kterým nalezneme její řešení, se nazývá

metoda LU rozkladu.

vyjádření neznámé ze vzorce.

zpětný chod.

přímý chod.

4. (1b.) Vyberte postup řešení soustavy $A\mathbf{x} = B$ metodou LU rozkladu.

$$A\mathbf{x} = B \rightarrow LU\mathbf{x} = B \rightarrow \mathbf{x} = L^{-1}U^{-1}B.$$

$$A\mathbf{x} = B \rightarrow LU\mathbf{x} = B \rightarrow U\mathbf{x} = L^{-1}B \rightarrow \mathbf{x} = L^{-1}BU^{-1}.$$

$$A\mathbf{x} = B \rightarrow LU\mathbf{x} = B \rightarrow L\mathbf{y} = B \rightarrow U\mathbf{x} = \mathbf{y}.$$

$$A\mathbf{x} = B \rightarrow LU\mathbf{x} = B \rightarrow U\mathbf{x} = B \rightarrow L\mathbf{x} = \mathbf{y}.$$

5. (1b.) Nechť reziduum $\mathbf{r} = B - A\tilde{\mathbf{x}}$ při řešení soustavy $A\mathbf{x} = B$ je malé. Kdy je malá relativní chyba řešení $\|\delta\mathbf{x}\|/\|\mathbf{x}\|$?

Vždycky.

Když $\kappa(A) \geq 100$.

Když $\|A\| \leq 1$.

Když je matice A dobře podmíněná.

6. (1b.) Vyberte, co platí pro přímé metody numerického řešení soustavy $A\mathbf{x} = B$.

Zkonstruujeme posloupnost vektorů $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots$, která konverguje k přesnému řešení.

Po konečném počtu kroků, když je splněna vhodná zastavovací podmínka, získáme přesné řešení.

Po konečném počtu kroků dostaneme řešení. Protože se nevyhneme zaokrouhlování, získané řešení bude jen přibližné.

Přímé metody používáme v případech, kdy máme velmi velký počet neznámých (řádově tisíce), protože jejich výpočtová náročnost je malá (rychleji konvergují k přesnému řešení).

7. (1b.) Na jakém principu jsou založeny iterační metody řešení soustavy $A\mathbf{x} = B$?

Výchozí soustavu $A\mathbf{x} = B$ převedeme pomocí řádkových elementárních úprav na ekvivalentní soustavu $C\mathbf{x} = D$.

Výchozí soustavu $A\mathbf{x} = B$ převedeme na ekvivalentní soustavu $LU\mathbf{x} = B$, kde $A = LU$ je LU rozklad matice A .

Výchozí soustavu $A\mathbf{x} = B$ přepíšeme na soustavu $\mathbf{x} = T\mathbf{x} + C$, která je ekvivalentní s původní soustavou a umožňuje iterování.

8. (1b.) Kdy dáváme přednost při řešení soustav lineárních rovnic metodě LU rozkladu před GEM?

Když máme řešit více soustav $A\mathbf{x} = B$ s různými pravými stranami B , ale stejnou maticí A .

Když máme řešit více soustav $A\mathbf{x} = B$ se stejnými maticemi B , ale různými maticemi A .

V případě, že si nejsme jisti, že GEM nezhavaruje.

Metodě LU rozkladu nedáváme před GEM přednost nikdy.

9. (1b.) Algoritmus pro řešení soustavy $Ax = B$ se nazývá numericky stabilní, když jím získané řešení \tilde{x} se málo liší od přesného řešení x .

jím získané řešení \tilde{x} vyhovuje soustavě $\tilde{A}\tilde{x} = \tilde{B}$, kde matice A se málo liší od matice \tilde{A} a sloupec B se málo liší od sloupce \tilde{B} .

při přesném počítání (bez zaokrouhlování) jím získané řešení \tilde{x} se málo liší od přesného řešení x .

po konečném počtu kroků dává přesné řešení.

10. (1b.) Která z následujících vlastností nezaručuje při řešení soustavy $Ax = B$ Gaussovou-Seidelovou metodou konvergenci?

A je ryze řádkově diagonálně dominantní.

A je ryze sloupcově diagonálně dominantní.

A má nenulové diagonální prvky.

A je symetrická pozitivně definitní.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Kapitola 4

Interpolace a aproximace funkcí

Cíle

Po prostudování této kapitoly budete schopni vysvětlit:

- co rozumíme aproximací funkce a jaké typy aproximací používáme,
- co je to interpolační polynom a jaké metody jeho nalezení známe,
- co je to Hermitův interpolační polynom,
- co rozumíme kubickými splajny, jaké jsou jejich druhy a jaké mají přednosti,
- kdy používáme metodu nejmenších čtverců a jaká je její podstata,
- jak postupujeme při vyrovnání polynomy metodou nejmenších čtverců.

Někdy je třeba „složitou“ funkci $f(x)$ nahradit „jednodušší“ funkcí $\varphi(x)$, která je funkci $f(x)$ „blízká“ a která se snadno matematicky zpracovává nebo modeluje na počítači. Typická je situace, kdy nemáme k dispozici vzorec funkce $f(x)$, ale pouze funkční hodnoty v nějaké konečné množině bodů.

Předpokládáme, že funkci $f(x)$ známe v $n + 1$ tzv. tabulkových (uzlových) bodech x_0, \dots, x_n nějakého intervalu (a, b) . Body nemusí být v tabulce seřazeny podle velikosti, ale musí být navzájem různé. Budeme používat označení $f(x_i) = y_i$ nebo $f(x_i) = f_i$. Tedy

x_i	x_0	x_1	\dots	x_{n-1}	x_n
y_i	y_0	y_1	\dots	y_{n-1}	y_n

Velmi často hledáme aproximující funkci $\varphi(x)$ jako lineární kombinaci několika jednoduchých základních funkcí

$$\varphi(x) = c_0\varphi_0(x) + \dots + c_m\varphi_m(x),$$

kde c_0, \dots, c_m jsou vhodné konstanty a $\varphi_0(x), \dots, \varphi_m(x)$ jsou pevně dané funkce. Nejčastější případy jsou:

1. $\varphi_0(x) = 1, \varphi_1(x) = x, \dots, \varphi_m(x) = x^m, \dots$; pak $\varphi(x)$ je polynom,
2. $\varphi_0(x) = 1, \varphi_1(x) = \cos x, \varphi_2(x) = \sin x, \varphi_3(x) = \cos 2x, \varphi_4(x) = \sin 2x, \dots$; pak je $\varphi(x)$ tzv. trigonometrický polynom, jehož použití je vhodné pro aproximaci 2π -periodických funkcí (případ obecné periody $2l$ lze vyřešit substitucí $x \mapsto (\pi/l)x$).

Označení. Pro uzavřený interval, jehož konci jsou nejmenší a největší z čísel x_0, \dots, x_n , použijeme označení $\text{Int}\langle x_0, \dots, x_n \rangle$, tj.

$$\text{Int}\langle x_0, \dots, x_n \rangle = \langle \min(x_0, \dots, x_n), \max(x_0, \dots, x_n) \rangle.$$

Obdobně pro otevřený interval použijeme označení $\text{Int}(x_0, \dots, x_n)$.

Funkci $\varphi(x)$ používáme k řešení řady důležitých úloh, mezi něž patří např.

1. *interpolace* — tj. výpočet hodnot $f(x)$ v neuzlových bodech v intervalu $\text{Int}\langle x_0, \dots, x_n \rangle$,
2. *extrapolace* — tj. výpočet hodnot $f(x)$ mimo interval $\text{Int}\langle x_0, \dots, x_n \rangle$,
3. výpočet hodnot derivace $f'(x)$, určitého integrálu $\int_a^b f(x) dx$ apod.

Budeme chtít, aby funkce $\varphi(x)$ splňovala určité požadavky, a bude nás zajímat, zda taková funkce $\varphi(x) = c_0\varphi_0(x) + \dots + c_m\varphi_m(x)$ existuje a je jediná. V dalším textu budeme studovat dva základní typy aproximace:

1. **Interpolační aproximace:** Budeme hledat takovou funkci $\varphi(x)$, pro kterou platí

$$\varphi(x_i) = y_i, \quad i = 0, 1, \dots, n.$$

Je tedy podstatné, aby graf interpolační funkce $\varphi(x)$ procházel body o souřadnicích $[x_i, y_i]$, $i = 0, \dots, n$. Tento požadavek je přirozený, pokud se domníváme, že hodnoty y_i jsou přesné.

Omezíme se přitom na případ polynomů, kde výsledky jsou jednodušší. Složitější otázkou trigonometrických polynomů se zabývat nebudeme (tato problematika úzce souvisí s diskrétní Fourierovou transformací, viz např. [52, str. 59]).

2. **Aproximace metodou nejmenších čtverců:** Budeme hledat takovou funkci $\varphi(x)$, pro kterou platí

$$\varphi(x_i) \doteq y_i, \quad i = 0, 1, \dots, n,$$

přičemž přibližná rovnost je určena tak, aby součet druhých mocnin odchylek $y_i - \varphi(x_i)$ byl minimální. Graf interpolační funkce $\varphi(x)$ tudíž nemusí přesně procházet body o souřadnicích $[x_i, y_i]$, $i = 0, \dots, n$. Tento požadavek je přirozený, pokud se domníváme, že hodnoty y_i nejsou přesné (např. jsou zatíženy chybami měření).

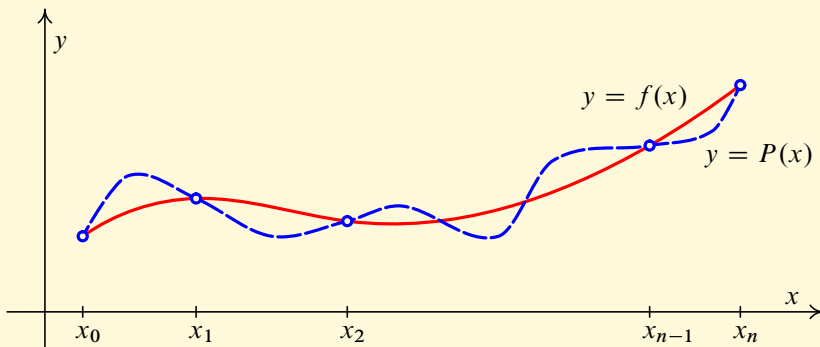
4.1 Interpolační polynom

Nechť je dáno $n + 1$ uzlových bodů x_0, x_1, \dots, x_n a jsou známy hodnoty funkce $f(x)$ v těchto bodech $y_0 = f(x_0), y_1 = f(x_1), \dots, y_n = f(x_n)$. Hledáme polynom $P(x)$ stupně nejvýše n ve tvaru

$$P(x) = c_0 + c_1x + \dots + c_nx^n, \quad (4.1)$$

pro nějž platí $P(x_i) = y_i$, $i = 0, \dots, n$. Jde tedy o interpolační aproximaci, kde volíme $\varphi_i(x) = x^i$, $i = 0, \dots, n$. Takový polynom budeme nazývat *interpolační polynom*. Tento polynom je znázorněn na obr. 4.1.

Polynom stupně n má $n + 1$ koeficientů, pro jejich určení máme $n + 1$ podmínek $P(x_i) = y_i$. Po jejich dosazení dostaneme obecně nehomogenní soustavu $n + 1$ lineárních



Obr. 4.1: Interpoláčn polynom $P(x)$ hodnot funkce $f(x)$

rovnice pro $n + 1$ neznámých c_0, \dots, c_n . Ve větě 4.1 ukážeme, že tato soustava má vždy jediné řešení.

4.1.1 Lagrangeův tvar interpolačního polynomu

Následující věta uvádí, že interpolační polynom existuje a je jediný. V jejím důkazu je současně popsán jeden z možných algoritmů jeho konstrukce (díky jednoznačnosti musí být výsledek vždy stejný). Výsledný tvar se nazývá *Lagrangeův¹ interpolační polynom*. Proto k jeho označení na rozdíl od vztahu (4.1) použijeme písmeno L .

Věta 4.1 *Pro libovolných $n + 1$ dvojic čísel $[x_i, y_i]$, $i = 0, \dots, n$, takových, že $x_i \neq x_j$ pro $i \neq j$, existuje právě jeden interpolační polynom $L(x)$.*

Důkaz.

1. *Existence.* Najdeme polynom s požadovanými vlastnostmi.

Nejdříve sestrojíme speciální tzv. *Lagrangeovy polynomy* $L_i(x)$, $i = 0, 1, \dots, n$, stupně n s následujícími vlastnostmi:

1. st $L_i(x) = n$.

¹**Joseph-Louis Lagrange** (1736–1813) (čti lagranž) — významný francouzský matematik italského původu. Zabýval se mechanikou, geometrií, diferenciálními rovnicemi, analýzou, algebrou, teorií čísel a dalšími matematickými obory a též teoretickou astronomií. Rozpracoval základní pojmy variačního počtu.

2. $L_i(x_i) = 1$, $L_i(x_j) = 0$ pro $i \neq j$.

Uvedeným požadavkům vyhovují polynomy $L_i(x)$ dané vztahy

$$L_i(x) = \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}, \quad i = 0, \dots, n. \quad (4.2)$$

Čitatel je totiž polynom, který nabývá nulových hodnot ve všech uzlech kromě x_i . V uzlu x_i pak nabývá nenulové hodnoty, která je stejná jako jmenovatel zlomku, takže platí $L_i(x_i) = 1$.

Nyní položíme

$$L(x) = y_0 L_0(x) + \cdots + y_n L_n(x). \quad (4.3)$$

Pro libovolné x_i platí:

$$\begin{aligned} L(x_i) &= y_0 L_0(x_i) + \cdots + y_i L_i(x_i) + \cdots + y_n L_n(x_i) = \\ &= y_0 \cdot 0 + \cdots + y_i \cdot 1 + \cdots + y_n \cdot 0 = y_i. \end{aligned}$$

Našli jsme tedy polynom s požadovanými vlastnostmi.

2. *Jednoznačnost.* Nechť také pro polynom $K(x)$, st $K \leq n$, platí $K(x_i) = y_i$, $i = 0, \dots, n$. Pak $M(x) = K(x) - L(x)$ je polynom stupně nanejvýš n nebo je nulový a $M(x_i) = K(x_i) - L(x_i) = y_i - y_i = 0$, pro $i = 0, \dots, n$. Tedy $M(x)$ má $n + 1$ kořenů, což znamená, že nutně $M(x)$ je nulový (nenulový polynom nemůže mít více kořenů, než je jeho stupeň), tj. $K(x) = L(x)$. \square

Poznámka 4.2

1. Pokud budou mít body $[x_0, y_0], \dots, [x_n, y_n]$ speciální polohu (např. budou ležet na přímce), může být stupeň $L(x)$ podstatně menší než n . Ale obecně je třeba polynom stupně n (tj. počet uzlových bodů snížený o jedničku).
2. Jestliže v předchozí větě připustíme i polynomy stupně většího než n , bude existovat nekonečně mnoho polynomů, které procházejí danými body $[x_0, y_0], \dots, [x_n, y_n]$. (Stačí přidat další body — jeden nebo více — s libovolnými souřadnicemi tak, aby aspoň jeden ležel mimo graf $L(x)$. Podle věty 4.1 bude existovat polynom, jehož graf prochází touto rozšířenou množinou bodů. Avšak jeho stupeň bude určitě vyšší než n .)

Příklad 4.3 Najděte Lagrangeův interpolační polynom pro data z následující tabulky.

x_i	-1	0	1	3
y_i	1	1	-1	2

Řešení. Podle postupu z věty 4.1 postupně dostaneme:

$$L_0(x) = \frac{(x-0)(x-1)(x-3)}{-1 \cdot (-2) \cdot (-4)} = -\frac{1}{8}x(x-1)(x-3),$$

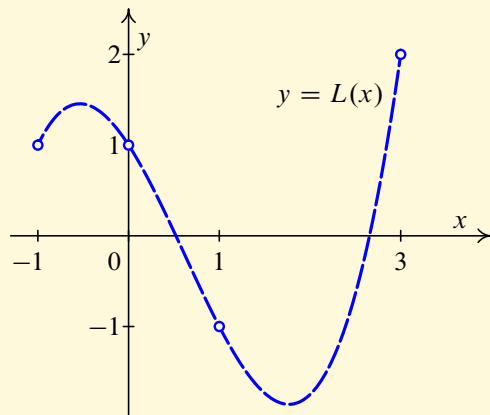
$$L_1(x) = \frac{(x+1)(x-1)(x-3)}{1 \cdot (-1) \cdot (-3)} = \frac{1}{3}(x+1)(x-1)(x-3),$$

$$L_2(x) = \frac{(x+1)(x-0)(x-3)}{2 \cdot 1 \cdot (-2)} = -\frac{1}{4}(x+1)x(x-3),$$

$$L_3(x) = \frac{(x+1)(x-0)(x-1)}{4 \cdot 3 \cdot 2} = \frac{1}{24}(x+1)x(x-1),$$

$$L(x) = 1 \cdot L_0(x) + 1 \cdot L_1(x) - 1 \cdot L_2(x) + 2 \cdot L_3(x) =$$

$$\begin{aligned}
 &= -\frac{1}{8}x(x-1)(x-3) + \\
 &\quad + \frac{1}{3}(x+1)(x-1)(x-3) + \\
 &\quad + \frac{1}{4}(x+1)x(x-3) + \\
 &\quad + \frac{1}{12}(x+1)x(x-1) = \\
 &= \frac{13}{24}x^3 - x^2 - \frac{37}{24}x + 1.
 \end{aligned}$$



Graf Lagrangeova polynomu je znázorněn na obr. 4.2.

▲ Obr. 4.2: $L(x) = \frac{13}{24}x^3 - x^2 - \frac{37}{24}x + 1$

Jak již bylo řečeno, interpolační polynom má mnoho důležitých aplikací, např. je základem pro odvození vzorců pro numerické derivování a integrování a řešení diferenciálních rovnic, jak uvidíme v následujících kapitolách.

Na závěr si všimneme, jak dobrou aproximací je interpolační polynom. Kvalita aproximace je obecně lepší uprostřed intervalů a horší u okrajů. Záleží na volbě uzlů, zejména v případě ekvidistantních uzlů je chyba u okrajů velká. Rozhodně je obvykle zcela nevhodný na extrapolaci. Pro chybu aproximace lze dokázat následující vztah (např. [3, str. 37], [22, str. 167], [43, str. 63] nebo [52, str. 49]).

Věta 4.4 *Nechť funkce $f(x)$ má derivaci řádu $n + 1$ na intervalu $\langle a, b \rangle$ a $L(x)$ je Lagrangeův interpolační polynom zkonstruovaný v bodech $x_0, \dots, x_n \in \langle a, b \rangle$. Označme*

$$f(x) = L(x) + R(x).$$

Pak pro libovolné $x \in \langle a, b \rangle$ lze chybu $R(x)$ vyjádřit ve tvaru

$$R(x) = (x - x_0)(x - x_1) \cdots (x - x_n) \cdot \frac{f^{(n+1)}(\xi)}{(n+1)!},$$

kde $\xi \in \text{Int}\langle x_0, \dots, x_n, x \rangle$ je vhodné číslo. Je-li tento interval nedegenerovaný, je ξ jeho vnitřní bod (k degeneraci intervalu dojde pro $n = 0$ a $x = x_0$, pak $\xi = x_0$).

Na velikost chyby má mimo jiného velký vliv polynom $(x - x_0) \cdots (x - x_n)$. Ukážeme, že při větším počtu uzlů mohou být chyby na okrajích intervalu skutečně veliké.

Příklad 4.5 Je dána funkce $f(x) = \frac{1}{x^2 + 1}$. Najděte a znázorněte Lagrangeův interpolační polynom, který prochází body $[x_i, f(x_i)]$, kde $x_i = -5, -4, \dots, 4, 5$.

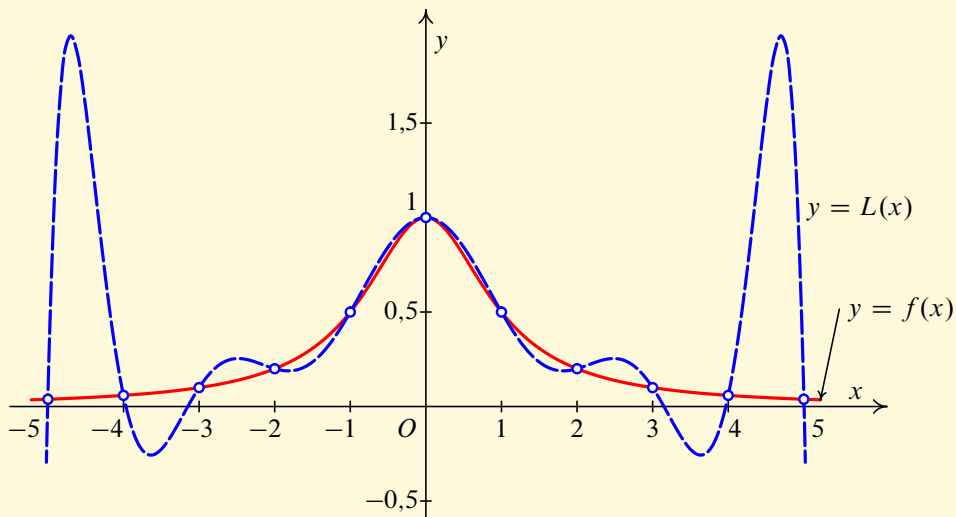
Řešení. Polynom $L(x)$ má procházet jedenácti body danými tabulkou

x_i	-5	-4	-3	-2	-1	0	1	2	3	4	5
y_i	$\frac{1}{26}$	$\frac{1}{17}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{17}$	$\frac{1}{26}$

Jeho stupeň tedy bude nejvýše deset. Podle věty 4.1 je $L(x) = y_0L_0(x) + \dots + y_{10}L_{10}$. Po úpravě vyjde

$$L(x) = 1 - \frac{149}{221}x^2 + \frac{2181}{11050}x^4 - \frac{83}{3400}x^6 + \frac{7}{5525}x^8 - \frac{1}{44200}x^{10}.$$

Z obrázku 4.3 je vidět, že na okrajích intervalu $\langle -5, 5 \rangle$ je aproximace funkce $f(x)$ interpolačním polynomem $L(x)$ velmi špatná, polynom zde osciluje. Pro rostoucí počet uzlů u této funkce chyba interpolace neomezeně roste (tzv. *Rungeho jev*). Podrobněji např. [42, str. 38], důkaz, který je netriviální, viz [27, str. 275] nebo [66]. ▲



Obr. 4.3: Graf funkce $f(x) = \frac{1}{x^2+1}$ a interpolačního polynomu $L(x)$

4.1.2 Newtonův tvar interpolačního polynomu

Algoritmus pro výpočet Lagrangeova interpolačního polynomu má jednu nevýhodu. Rozhodneme-li se dodatečně přidat jeden uzlový bod, musíme všechny výpočty provést znovu. Existují postupy založené na tzv. *iterované interpolaci*, kdy se uzly postupně přidávají, které tento problém odstraňují. Nejznámější jsou *Nevillův¹ algoritmus* a *Aitkenův² algoritmus* (např. [22, str. 181] nebo [52, str. 40]). Nicméně tyto metody jsou vhodné pro výpočet hodnoty interpolačního polynomu v jednom konkrétním bodě, pro nalezení obecného vzorce se příliš nehodí. Proto si v tomto oddíle popíšeme jiný způsob nalezení interpolačního polynomu.

Předpokládejme opět, že je dáno $n + 1$ uzlových bodů x_0, x_1, \dots, x_n a hodnoty funkce $f(x)$ v těchto bodech $f(x_i) = y_i, i = 0, \dots, n$. Interpolační polynom $N(x)$ budeme hledat ve tvaru

$$N(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1}). \quad (4.4)$$

¹**Eric Harold Neville** (1889–1961) (čti nevil) — anglický matematik.

²**Alexander Craig Aitken** (1895–1967) (čti eitken) — novozélandský matematik.

Dosazením se snadno ověří, že takové konstanty a_i existují a jsou určeny jednoznačně. Stačí postupně dosazovat hodnoty x_0, \dots, x_n :

$$N(x_0) = y_0 \quad \Rightarrow \quad a_0 = y_0,$$

$$N(x_1) = y_1 \quad \Rightarrow \quad a_0 + a_1(x_1 - x_0) = y_1 \quad \Rightarrow \quad a_1 = \frac{y_1 - y_0}{x_1 - x_0},$$

$$N(x_2) = y_2 \quad \Rightarrow \quad a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) = y_2$$

$$\Rightarrow \quad a_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_0} \text{ atd.}$$

Polynom $N(x)$ se nazývá *Newtonův¹ interpolační polynom*. Z jednoznačnosti interpolačního polynomu (věta 4.1) plyne, že nutně $L(x) = N(x)$, kde $L(x)$ je Lagrangeův interpolační polynom pro daná data. Tedy je to jen jiný zápis Lagrangeova interpolačního polynomu! Správnější je tedy mluvit o Lagrangeově nebo Newtonově tvaru interpolačního polynomu.

¹**Isaac Newton** (1643–1727) (čti njútn) — anglický matematik, fyzik, mechanik a astronom. Položil základy diferenciálního a integrálního počtu, který potřeboval pro vybudování klasické mechaniky.

Ukážeme si, jak lze koeficienty a_k Newtonova polynomu snadno nalézt. K tomu zavedeme tzv. *poměrné difference*.

Poměrné difference

Poměrné difference jsou jakousi náhradou derivací pro funkce, které jsou definované pouze na diskrétní množině bodů x_0, \dots, x_n . Připomeňme, že tyto body nemusí být uspořádané podle velikosti, ale musí být navzájem různé.

Poměrné difference k -tého řádu, kde $k = 0, 1, \dots, n$, definujeme rekurentně takto:

$$\text{Nultý řád: } f[x_i] = f(x_i), \quad i = 0, \dots, n.$$

$$\text{První řád: } f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i}, \quad i = 0, \dots, n - 1.$$

$$k\text{-tý řád: } f[x_i, x_{i+1}, \dots, x_{i+k}] = \frac{f[x_{i+1}, \dots, x_{i+k}] - f[x_i, \dots, x_{i+k-1}]}{x_{i+k} - x_i},$$

$$i + k \leq n.$$

Poměrné diference $f[x_i, x_{i+1}, \dots, x_{i+k}]$ je možné vyjádřit pomocí hodnot y_i, y_{i+1} až y_{i+k} . Indukcí lze ukázat (viz např. [3, str. 38]), že

$$\begin{aligned} f[x_i, x_{i+1}, \dots, x_{i+k}] &= \sum_{j=0}^k \frac{y_{i+j}}{\prod_{\substack{r=0 \\ r \neq j}}^k (x_{i+j} - x_{i+r})} = \\ &= \sum_{j=0}^k \frac{y_{i+j}}{(x_{i+j} - x_i) \cdots (x_{i+j} - x_{i+j-1})(x_{i+j} - x_{i+j+1}) \cdots (x_{i+j} - x_{i+k})}. \end{aligned}$$

Z tohoto vztahu vyplývá, že *poměrné diference jsou symetrické vzhledem ke svým argumentům*, tedy hodnoty $f[x_i, x_{i+1}, \dots, x_{i+k}]$ nezávisí na pořadí uzlů, v nichž se počítají. Lze dokázat, že pro koeficienty polynomu $N(x)$ platí $a_k = f[x_0, \dots, x_k]$, kde $k = 0, \dots, n$ (viz [3, str. 39], [52, str. 44]). Tedy Newtonův interpolační polynom má tvar

$$\begin{aligned} N(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \\ &\quad + \cdots + f[x_0, \dots, x_n](x - x_0) \cdots (x - x_{n-1}). \end{aligned} \tag{4.5}$$

Výpočet poměrných diferencí je vhodné zorganizovat do následující tabulky (červeně vyznačené poměrné diference budeme potřebovat do Newtonova interpolačního polynomu):

i	x_i	$f[x_i]$	$f[x_{i-1}, x_i]$	\dots	$f[x_{i-n+1}, \dots, x_i]$	$f[x_{i-n}, \dots, x_i]$
0	x_0	$f[x_0]$				
1	x_1	$f[x_1]$	\rightarrow	$f[x_0, x_1]$		
\vdots	\vdots		\vdots		\ddots	
$n-1$	x_{n-1}	$f[x_{n-1}]$	\rightarrow	$f[x_{n-2}, x_{n-1}]$	\dots	\rightarrow
n	x_n	$f[x_n]$	\rightarrow	$f[x_{n-1}, x_n]$	\dots	\rightarrow
						$f[x_0, \dots, x_n]$

Tab. 4.1: Výpočet poměrných diferencí

Příklad 4.6 Jsou dány uzly $x_0 = -1$, $x_1 = 0$, $x_2 = 1$, $x_3 = 3$ a funkční hodnoty $f(x_0) = 1$, $f(x_1) = 1$, $f(x_2) = -1$, $f(x_3) = 2$. Najděte pro ně Newtonův interpolační polynom.

Řešení. Potřebujeme vypočítat poměrné diference. Výpočet uspořádáme do tabulky:

i	x_i	$y_i = f(x_i) = f[x_i]$	$f[x_{i-1}, x_i]$	$f[x_{i-2}, x_{i-1}, x_i]$	$f[x_{i-3}, x_{i-2}, x_{i-1}, x_i]$
0	-1	$f[x_0] = 1$			
1	0	$f[x_1] = 1$	$f[x_0, x_1] = \frac{1-1}{0-(-1)} = 0$		
2	1	$f[x_2] = -1$	$f[x_1, x_2] = \frac{-1-1}{1-0} = -2$	$f[x_0, x_1, x_2] = \frac{-2-0}{1-(-1)} = -1$	
3	3	$f[x_3] = 2$	$f[x_2, x_3] = \frac{2-(-1)}{3-1} = \frac{3}{2}$	$f[x_1, x_2, x_3] = \frac{\frac{3}{2}-(-2)}{3-0} = \frac{7}{6}$	$f[x_0, x_1, x_2, x_3] = \frac{\frac{7}{6}-(-1)}{3-(-1)} = \frac{13}{24}$

Podle vzorce (4.5) tedy platí

$$\begin{aligned}
 N(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \\
 &\quad + f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) = \\
 &= 1 + 0 \cdot (x + 1) - 1 \cdot (x + 1)x + \frac{13}{24} (x + 1)x(x - 1) = \\
 &= 1 - (x + 1)x + \frac{13}{24} (x + 1)x(x - 1) = \\
 &= 1 + (x + 1) \left(0 + x \left(-1 + \frac{13}{24} (x - 1) \right) \right).
 \end{aligned}$$

Zápis s postupným vytknutím je stručnější a efektivnější při dosazování za konkrétní x (dosazuje se zprava a zevnitř, jde vlastně o obdobu Hornerova schématu).

Protože data ze zadání tohoto příkladu jsou stejná jako v příkladu 4.3, musí jít díky jednoznačnosti interpolačního polynomu o tentýž polynom, rozdíl je jen v zápisu. Zejména graf na obr. 4.2 je současně i grafem vypočteného Newtonova polynomu $N(x)$. ▲

U Newtonova interpolačního polynomu je jednoduché přidat další uzel. Protože uzly nemusí být uspořádané podle velikosti, stačí přidat do tabulky dolů jeden řádek s novým uzlem, připsat doprava jeden sloupec a dopočítat chybějící poměrné diference. Obdobně můžeme snadno přidat další uzly.

4.1.3 Newtonův interpolační polynom — případ ekvidistantních uzlů

Všimneme si, jak se Newtonův tvar interpolačního polynomu zjednoduší, když budou uzly ekvidistantní a uspořádané podle velikosti. Budeme tedy předpokládat, že $x_1 = x_0 + h$, $x_2 = x_0 + 2h, \dots, x_n = x_0 + nh$, kde $h > 0$ je konstantní *krok*. Aby se zápis co nejméně zjednodušil, je účelné zavést obdobu poměrných diferencí pro ekvidistantní uzly. Jedná se opět o jakousi diskrétní obdobu derivací. Definice jsou zase rekurentní.

1) *Diference vpřed* neboli *dopředné diference*:

$$\Delta^0 y_i = y_i, i = 0, \dots, n, \dots \text{ nultá diference vpřed,}$$

$$\Delta y_i = \Delta^1 y_i = y_{i+1} - y_i, i = 0, \dots, n - 1, \dots \text{ první diference vpřed,}$$

$$\begin{aligned} \Delta^2 y_i &= \Delta(\Delta y_i) = \Delta(y_{i+1} - y_i) = (y_{i+2} - y_{i+1}) - (y_{i+1} - y_i) = \\ &= y_{i+2} - 2y_{i+1} + y_i, i = 0, \dots, n - 2, \dots \text{ druhá diference vpřed atd.} \end{aligned}$$

Obecně definujeme *k-tou diferenci vpřed* neboli *diferenci vpřed k-tého řádu* vztahem

$$\Delta^k y_i = \Delta(\Delta^{k-1} y_i) = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i, \quad i = 0, \dots, n - k.$$

Indukcí lze ukázat, že platí vzorec

$$\begin{aligned}\Delta^k y_i &= \sum_{j=0}^k (-1)^j \binom{k}{j} y_{i+k-j} = \\ &= y_{i+k} - \binom{k}{1} y_{i+k-1} + \binom{k}{2} y_{i+k-2} - \cdots (-1)^k y_i,\end{aligned}$$

kde $\binom{m}{n} = \frac{m!}{(m-n)!n!}$ pro $m, n \in \mathbb{Z}_0^+$, $m \geq n$, je kombinační číslo. Klademe $0! = 1$.

2) *Diference vzad* neboli *zpětné diference*:

$\nabla^0 y_i = y_i, i = 0, \dots, n, \dots$ nulová diference vzad,

$\nabla y_i = \nabla^1 y_i = y_i - y_{i-1}, i = 1, \dots, n, \dots$ první diference vzad,

$\nabla^2 y_i = \nabla(\nabla y_i) = \nabla(y_i - y_{i-1}) = (y_i - y_{i-1}) - (y_{i-1} - y_{i-2}) =$
 $= y_i - 2y_{i-1} + y_{i-2}, i = 2, \dots, n, \dots$ druhá diference vzad atd.

Obecně definujeme *k-tou diferenci vzad* neboli *diferenci vzad k-tého řádu* vztahem

$$\nabla^k y_i = \nabla(\nabla^{k-1} y_i) = \nabla^{k-1} y_i - \nabla^{k-1} y_{i-1}, \quad i = k, \dots, n.$$

Indukcí lze ukázat, že platí vzorec

$$\nabla^k y_i = \sum_{j=0}^k (-1)^j \binom{k}{j} y_{i-j} = y_i - \binom{k}{1} y_{i-1} + \binom{k}{2} y_{i-2} - \cdots (-1)^k y_{i-k}.$$

Z předchozího vzorce a obdobného vzorce pro diference vpřed je vidět, že

$$\nabla^k y_i = \Delta^k y_{i-k} \quad \text{resp.} \quad \Delta^k y_i = \nabla^k y_{i+k}. \quad (4.6)$$

Výpočet diferencí vpřed a vzad je výhodné zapsat do následujících tabulek.

i	x_i	y_i	Δy_i	...	$\Delta^{n-1} y_i$	$\Delta^n y_i$
0	x_0	y_0	$\rightarrow \Delta y_0$...	$\rightarrow \Delta^{n-1} y_0$	$\rightarrow \Delta^n y_0$
1	x_1	y_1	$\rightarrow \Delta y_1$...	$\rightarrow \Delta^{n-1} y_1$	
\vdots	\vdots	\vdots	\vdots	\ddots		
$n-1$	x_{n-1}	y_{n-1}	$\rightarrow \Delta y_{n-1}$			
n	x_n	y_n				

i	x_i	y_i	∇y_i	...	$\nabla^{n-1} y_i$	$\nabla^n y_i$
0	x_0	y_0				
1	x_1	y_1	$\rightarrow \nabla y_1$			
\vdots	\vdots	\vdots	\vdots	\ddots		
$n-1$	x_{n-1}	y_{n-1}	$\rightarrow \nabla y_{n-1}$...	$\rightarrow \nabla^{n-1} y_{n-1}$	
n	x_n	y_n	$\rightarrow \nabla y_n$...	$\rightarrow \nabla^{n-1} y_n$	$\rightarrow \nabla^n y_n$

Tab. 4.2: Výpočet diferencí vpřed a vzad

Vzhledem ke vztahům (4.6) obsahují obě tabulky tytéž hodnoty a jsou tedy rovnocenné. Zejména na vedlejší diagonále první tabulky najdeme hodnoty $y_n, \nabla y_n, \nabla^2 y_n, \dots, \nabla^n y_n$ a na hlavní diagonále druhé tabulky najdeme hodnoty $y_0, \Delta y_0, \Delta^2 y_0, \dots, \Delta^n y_0$.

Nyní se vrátíme k Newtonovu polynomu. Budeme uvažovat dva případy — uzly seřazené vzestupně a sestupně. Nejprve najdeme vztah mezi poměrnými a dopřednými a zpětnými diferencemi. Platí

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{y_{i+1} - y_i}{h} = \frac{\Delta y_i}{h},$$

$$f[x_i, x_{i-1}] = \frac{f(x_{i-1}) - f(x_i)}{x_{i-1} - x_i} = \frac{y_{i-1} - y_i}{-h} = \frac{-\nabla y_i}{-h} = \frac{\nabla y_i}{h}.$$

Z těchto vztahů se indukcí dokáže, že obecně platí

$$f[x_i, x_{i+1}, \dots, x_{i+k}] = \frac{\Delta^k y_i}{k! h^k}, \quad f[x_i, x_{i-1}, \dots, x_{i-k}] = \frac{\nabla^k y_i}{k! h^k}. \quad (4.7)$$

Abychom zjednodušili označení, zavedeme si ještě tzv. *zobecněný binomický koeficient*.

Pro $t \in \mathbb{R}$ a $j \in \mathbb{N}$ definujeme

$$\binom{t}{j} = \frac{t(t-1)\cdots(t-j+1)}{j!}; \quad \text{pro } j = 0 \text{ klademe } \binom{t}{0} = 1.$$

Pro $t \in \mathbb{N}$, $0 \leq j \leq t$, je tato definice ve shodě s obyčejným binomickým koeficientem.

(I) Uzly seřadíme vzestupně, tj. $x_0 < x_1 < \cdots < x_n$. Zavedeme substituci $x = x_0 + th$. Odtud dostaneme

$$t = \frac{x - x_0}{h}, \quad \frac{x - x_i}{h} = \frac{x_0 + th - (x_0 + ih)}{h} = t - i, \quad i = 0, \dots, n.$$

Po dosazení do (4.5) s využitím předchozích vztahů po úpravě vyjde

$$\begin{aligned} N^+(t) &= N(x) = & (4.8) \\ &= y_0 + \frac{\Delta y_0}{1!} t + \frac{\Delta^2 y_0}{2!} t(t-1) + \cdots + \frac{\Delta^n y_0}{n!} t(t-1)\cdots(t-n+1) = \\ &= y_0 + \binom{t}{1} \Delta y_0 + \binom{t}{2} \Delta^2 y_0 + \cdots + \binom{t}{n} \Delta^n y_0 = \sum_{j=0}^n \binom{t}{j} \Delta^j y_0. \end{aligned}$$

Polynom $N^+(t)$ se nazývá *Newtonův interpolační polynom vpřed*.

(II) Uzly seřadíme sestupně, tj. $x_n > x_{n-1} > \dots > x_0$. Zavedeme substituci $x = x_n + sh$. Odtud dostaneme

$$s = \frac{x - x_n}{h}, \quad \frac{x - x_{n-i}}{h} = \frac{x_n + sh - (x_n - ih)}{h} = s + i, \quad i = 0, \dots, n.$$

Po dosazení do (4.5), kde zaměníme x_0, x_1, \dots, x_n za x_n, x_{n-1}, \dots, x_0 , s využitím předchozích vztahů po úpravě vyjde

$$\begin{aligned} N^-(s) &= N(x) = & (4.9) \\ &= y_n + \frac{\nabla y_n}{1!} s + \frac{\nabla^2 y_n}{2!} s(s+1) + \dots + \frac{\nabla^n y_n}{n!} s(s+1)\dots(s+n-1) = \\ &= y_n - \binom{-s}{1} \nabla y_n + \binom{-s}{2} \nabla^2 y_n - \dots + (-1)^n \binom{-s}{n} \nabla^n y_n = \\ &= \sum_{j=0}^n (-1)^j \binom{-s}{j} \nabla^j y_n. \end{aligned}$$

Polynom $N^-(s)$ se nazývá *Newtonův interpolační polynom vzad*.

Často se setkáváme se situací, kdy máme rozsáhlou tabulku s hodnotami $[x_i, y_i]$ odpovídajícími neznámé funkci f a potřebujeme určit hodnotu funkce f v bodě \bar{x} , který není uzlový a leží mezi x_0 a x_n . Pak můžeme nahradit funkci f interpolačním polynomem L a položit $f(\bar{x}) \approx L(\bar{x})$. Jak již víme, použijeme-li na konstrukci interpolačního polynomu příliš mnoho uzlových bodů, dostaneme obecně špatnou aproximaci. Proto vybereme jen několik uzlových bodů, které jsou blízko čísla \bar{x} , a ty použijeme ke konstrukci interpolačního polynomu. Jsou-li uzly ekvidistantní, lze s výhodou použít Newtonův interpolační polynom vpřed nebo vzad. Ty jsou zvlášť vhodné, jestliže číslo \bar{x} leží na začátku nebo konci tabulky. Pokud se totiž rozhodneme zvýšit počet použitých uzlů, snadno je přidáme. Tyto polynomy jsou vhodné pro nalezení jedné konkrétní funkční hodnoty, nehodí se příliš pro nalezení obecného vzorce interpolačního polynomu.

Příklad 4.7 V následující tabulce jsou dány hodnoty neznámé funkce f :

x_i	0,0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8	2,0
y_i	0,96	0,94	0,90	0,79	0,59	0,27	-0,17	-0,64	-0,96	-0,92	-0,40

Pomocí vhodného interpolačního polynomu určete přibližně $f(0,3)$ a $f(1,7)$.

Řešení. Zvolíme např. interpolační polynomy třetího stupně a vybereme čtyři uzlové body ze začátku resp. z konce tabulky. Protože uzly jsou ekvidistantní s krokem $h = 0,2$, můžeme použít Newtonův interpolační polynom vpřed resp. vzad. Nejprve vypočítáme potřebné diference vpřed resp. vzad. Výpočty uspořádáme do tabulek.

Tabulka diferencí vpřed pro první čtyři uzly je:

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$
0	0	0,96	-0,02	-0,02	-0,05
1	0,2	0,94	-0,04	-0,07	
2	0,4	0,90	-0,11		
3	0,6	0,79			

Odpovídající Newtonův interpolační polynom vpřed je:

$$\begin{aligned} N^+(t) &= y_0 + \frac{\Delta y_0}{1!} t + \frac{\Delta^2 y_0}{2!} t(t-1) + \frac{\Delta^3 y_0}{3!} t(t-1)(t-2) = \\ &= 0,96 - 0,02t - 0,01t(t-1) - (0,05/6)t(t-1)(t-2). \end{aligned}$$

Pro $t = \frac{x-x_0}{h} = \frac{0,3-0}{0,2} = 1,5$ vyjde $N^+(1,5) = 0,925\ 625$, tedy $f(0,3) \approx 0,925\ 625$.

Tabulka diferencí vzad pro poslední čtyři uzly (přeznačené na x_0 až x_3) je:

i	x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$
0	1,4	-0,64			
1	1,6	-0,96	-0,32		
2	1,8	-0,92	0,04	0,36	
3	2	-0,40	0,52	0,48	0,12

Odpovídající Newtonův interpolační polynom vzad je:

$$\begin{aligned} N^-(s) &= y_3 + \frac{\nabla y_3}{1!} s + \frac{\nabla^2 y_3}{2!} s(s+1) + \frac{\nabla^3 y_3}{3!} s(s+1)(s+2) = \\ &= -0,40 + 0,52s + 0,24s(s+1) + 0,02s(s+1)(s+2). \end{aligned}$$

Pro $s = \frac{x-x_3}{h} = \frac{1,7-2}{0,2} = -1,5$ vyjde $N^-(-1,5) = -0,992\,500$, tedy platí $f(1,7) \approx \approx -0,992\,500$.

Hodnoty v zadané tabulce jsou hodnoty funkce $g(x) = \cos(x^2 + 0,3)$ zaokrouhlené na dvě desetinná místa. Přesné hodnoty jsou tudíž pro porovnání $g(0,3) \doteq 0,924\,909$ a $g(1,7) \doteq -0,998\,829$ (zaokrouhleno na šest platných cifer). ▲

Poznámka 4.8 Různými způsoby uspořádání uzlů lze získat další interpolační polynomy. Mezi ně patří např. Gaussovy interpolační vzorce, interpolační vzorec Stirlingův¹, Besselův², Everettův³

¹**James Stirling** (1692–1770) — skotský matematik.

²**Friedrich Wilhelm Bessel** (1784–1846) — německý matematik a fyzik.

³**Joseph David Everett** (1831–1904) — anglický fyzik.

a Steffensenův¹ — viz [3, 4, 29, 43, 51]. Některé jsou vhodné pro nalezení hodnoty v bodě, který leží uprostřed tabulky známých hodnot s ekvidistantními uzly, protože umožňují snadno přidávat další uzlové body z obou stran. Je však třeba připomenout, že vzhledem k jednoznačnosti interpolačního polynomu (věta 4.1) jde (při stejné množině uzlových bodů a hodnot v nich) jen o různé zápisy téhož polynomu.

4.1.4 Hermitův interpolační polynom

Lagrangeův interpolační polynom nabývá v daných uzlových bodech x_0, x_1, \dots, x_n předepsané hodnoty. Obecněji lze hledat polynom, který má v uzlových bodech předepsané i hodnoty některých derivací.

Nechť m_0, m_1, \dots, m_n jsou daná přirozená čísla a pro každé m_i jsou předepsána reálná čísla $y_i^{(j)}$, $j = 0, \dots, m_i - 1$, $i = 0, \dots, n$. Hledejme polynom $P_k(x)$ co nejnižšího stupně k takový, že pro něj platí $P_k^{(j)}(x_i) = y_i^{(j)}$,² $i = 0, \dots, n$, $j = 0, \dots, m_i - 1$ (symbol $P_k^{(j)}$ značí j -tou derivaci; přitom klademe $P_k^{(0)} = P_k$).

¹Johan Frederik Steffensen (1873–1961) — dánský matematik a statistik.

²U $P_k^{(j)}$ horní index (j) značí j -tou derivaci, ne však u $y_i^{(j)}$, kde slouží pouze k odlišení různých čísel.

Pro hledaný polynom je tedy předepsáno $m = m_0 + m_1 + \dots + m_n$ podmínek pro funkční hodnoty a hodnoty derivací v uzlových bodech (v i -tém uzlu je předepsáno m_i podmínek pro funkční hodnotu, hodnotu první derivace atd. až hodnotu derivace řádu $m_i - 1$). Protože polynomy stupně nejvýše $m - 1$ mají m koeficientů, což se shoduje s počtem uvedených podmínek, lze očekávat, že pro nejnižší stupeň k hledaného polynomu by mohlo platit $k = m - 1$. Tak tomu skutečně je. Následující věta obsahuje tvrzení o existenci a jednoznačnosti tohoto polynomu.

Věta 4.9 *Nechť x_0, x_1, \dots, x_n , kde $n \in \mathbb{N}$, jsou navzájem různá reálná čísla. Nechť m_0, m_1, \dots, m_n jsou přirozená čísla. Označme $m = m_0 + m_1 + \dots + m_n$. Dále necht' je dáno m reálných čísel $y_i^{(j)}$, $i = 0, \dots, n$, $j = 0, \dots, m_i - 1$.*

Pak mezi všemi polynomy stupně nejvýše $m - 1$ existuje právě jeden polynom $P(x)$, pro nějž platí $P^{(j)}(x_i) = y_i^{(j)}$, $i = 0, \dots, n$, $j = 0, \dots, m_i - 1$.

Důkaz. Označme hledaný polynom $P(x) = c_0 + c_1x + \dots + c_kx^k$, kde $k = m - 1$. Dosadíme-li do něj uzlový bod x_i , má platit $c_0 + c_1x_i + \dots + c_kx_i^k = y_i^{(0)}$, což je lineární rovnice pro neznámé koeficienty c_0, c_1, \dots, c_k . Derivace polynomu jsou opět polynomy, takže obdobně za každou předepsanou hodnotu derivace v některém uzlovém bodě dostaneme lineární rovnici pro neznámé koeficienty.

Protože počet neznámých koeficientů i počet předepsaných podmínek je roven číslu m , dostaneme čtvercovou soustavu lineárních rovnic. Označme její matici soustavy A , sloupec pravých stran B a sloupec neznámých c . Zřejmě složky sloupce B jsou předepsané funkční hodnoty a hodnoty derivací $y_i^{(j)}$. Soustavu pak můžeme stručně zapsat v maticovém tvaru $Ac = B$.

Ukážeme, že matice A je regulární. To je rovnocenné tomu, že homogenní čtvercová soustava $Ac = O$, kde O je nulový sloupec, má jediné řešení.

Připomeňme poznatek z algebry, že číslo α (reálné nebo komplexní) je aspoň r -násobným kořenem polynomu $Q(x)$ právě tehdy, když platí $Q(\alpha) = Q'(\alpha) = \dots = Q^{(r-1)}(\alpha) = 0$. (Je-li $Q^{(r)}(\alpha) \neq 0$, je α právě r -násobný kořen.)

Volba $B = O$ odpovídá tomu, že všechny předepsané funkční hodnoty i hodnoty derivací jsou nulové. Tudíž všechny uzlové body x_i jsou kořeny, jejichž násobnosti jsou alespoň m_i . Celkový počet kořenů včetně násobností je proto aspoň $m_0 + \dots + m_n = m$.

Homogenní soustava $Ac = O$ má vždy triviální řešení $c_0 = \dots = c_k = 0$, které odpovídá nulovému polynomu. Připusťme, že existuje i nějaké netriviální řešení. To by znamenalo, že existuje nenulový polynom stupně nejvýše $m - 1$, který má aspoň m kořenů, což je spor.

Protože je matice A regulární, má soustava $Ac = B$ pro libovolný sloupec pravých stran B , tj. pro libovolné předepsané funkční hodnoty a hodnoty derivací $y_i^{(j)}$ právě jedno řešení, což jsme měli dokázat. \square

Polynom z předchozí věty se nazývá *Hermitův¹ interpolační polynom*. V případě $m_0 = \dots = m_n = 1$ přejde v Lagrangeův interpolační polynom, je tedy jeho zobecněním. Podobně jako Lagrangeův polynom může mít ve speciálním případě menší stupeň než je n , i Hermitův polynom může mít menší stupeň než je $m - 1$ (srovnejte poznámku 4.2).

Nalezení Hermitova interpolačního polynomu

Popíšeme tři způsoby nalezení Hermitova interpolačního polynomu.

- *Metoda neurčitých koeficientů* — půjde o přímé použití postupu z důkazu věty 4.9.
- *Obdoba vzorce (4.3) pro Lagrangeův interpolační polynom* — polynomy $L_i(x)$ nahradíme jejich zobecněními (viz např. [4, I. díl, str. 164] nebo [52, str. 52]). Kromě tohoto rekurentního vzorce existuje i uzavřený vzorec — viz věta 4.13.
- *Obdoba vzorce (4.5) pro Newtonův interpolační polynom* — poměrné diference nahradíme jejich zobecněními (viz např. [52, str. 56]).

¹**Charles Hermite** (1822–1901) (čti ermit) — významný francouzský matematik. Zabýval se matematickou analýzou, algebrou a teorií čísel.

Metoda neurčitých koeficientů

Při označení z věty 4.9 a jejího důkazu dostaneme pro koeficienty hledaného Hermitova polynomu $P(x) = c_0 + c_1x + \dots + c_{m-1}x^{m-1}$ soustavu m lineárních rovnic o m neznámých c_0, \dots, c_{m-1} , která má jediné řešení. Jednotlivé rovnice dostaneme, když do očekávaného vzorce pro $P(x)$ resp. jeho derivace dosadíme uzlové body a předepsané funkční hodnoty resp. hodnoty derivací. Postup si ukážeme na jednoduchém příkladu.

Příklad 4.10 Najděte Hermitův interpolační polynom $P(x)$, pro nějž platí $P(-1) = 2$, $P(1) = 0$, $P'(1) = -1$ a $P(2) = 1$.

Řešení. Při označení z věty 4.9 máme $n = 2$, $x_0 = -1$, $x_1 = 1$ a $x_2 = 2$. Protože v prvním a posledním uzlu jsou předepsané jen funkční hodnoty, kdežto v prostředím je předepsaná funkční hodnota a hodnota první derivace, je $m_0 = 1$, $m_1 = 2$ a $m_2 = 1$. Dále $y_0^{(0)} = 2$, $y_1^{(0)} = 0$, $y_1^{(1)} = -1$ a $y_2^{(0)} = 1$.

Jelikož jsou předepsány čtyři podmínky ($m = m_0 + m_1 + m_2 = 4$), má hledaný polynom nejvýše stupeň $m - 1 = 3$. Tedy

$$P(x) = c_0 + c_1x + c_2x^2 + c_3x^3.$$

Dále platí

$$P'(x) = c_1 + 2c_2x + 3c_3x^2.$$

Dosazením za uzlové body obdržíme z předepsaných podmínek následující soustavu lineárních rovnic:

$$\begin{aligned} c_0 - c_1 + c_2 - c_3 &= 2, \\ c_0 + c_1 + c_2 + c_3 &= 0, \\ c_1 + 2c_2 + 3c_3 &= -1, \\ c_0 + 2c_1 + 4c_2 + 8c_3 &= 1. \end{aligned}$$

Soustavu vyřešíme Gaussovou eliminační metodou. Přímý chod je:

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 2 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & -1 \\ 1 & 2 & 4 & 8 & 1 \end{array} \right) \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 2 \\ 0 & 2 & 0 & 2 & -2 \\ 0 & 1 & 2 & 3 & -1 \\ 0 & 3 & 3 & 9 & -1 \end{array} \right) \begin{array}{l} (2) - (1) \\ (4) - (1) \end{array} \sim$$

$$\begin{aligned}
& \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 3 & -1 \\ 0 & 2 & 0 & 2 & -2 \\ 0 & 3 & 3 & 9 & -1 \end{array} \right) \begin{array}{l} (3) \\ (2) \end{array} \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 3 & -1 \\ 0 & 0 & -4 & -4 & 0 \\ 0 & 0 & -3 & 0 & 2 \end{array} \right) \begin{array}{l} (3) - 2(2) \\ (4) - 3(2) \end{array} \sim \\
& \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 3 & -1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -3 & 0 & 2 \end{array} \right) -\frac{1}{4}(3) \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 3 & -1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 & 2 \end{array} \right) (4) + 3(3)
\end{aligned}$$

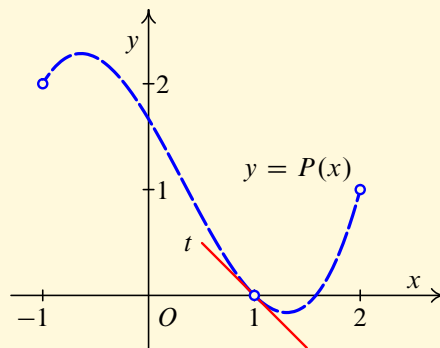
Zpětným chodem dostaneme:

$$\begin{array}{rcl}
c_0 - c_1 + c_2 - c_3 = 2, & & c_3 = \frac{2}{3}, \\
c_1 + 2c_2 + 3c_3 = -1, & \text{odkud} & c_2 = -c_3 = -\frac{2}{3}, \\
c_2 + c_3 = 0, & & c_1 = -1 - 2c_2 - 3c_3 = -\frac{5}{3}, \\
3c_3 = 2, & & c_0 = 2 + c_1 - c_2 + c_3 = \frac{5}{3}.
\end{array}$$

Hledaný Hermitův interpolační polynom je tedy

$$P(x) = \frac{5}{3} - \frac{5}{3}x - \frac{2}{3}x^2 + \frac{2}{3}x^3.$$

Jeho graf je znázorněn na obr. 4.4. Je vidět, že prochází třemi body $[-1, 2]$, $[1, 0]$ a $[2, 1]$. Navíc tečna t k tomuto grafu v bodě $[1, 0]$ má směrnici rovnou číslu -1 , protože platí rovnost $P'(0) = -1$. Její rovnice je tudíž $y - 0 = -1(x - 1)$, tj. $y = -x + 1$. ▲



Obr. 4.4: $P(x) = \frac{2}{3}x^3 - \frac{2}{3}x^2 - \frac{5}{3}x + \frac{5}{3}$

Poznámka 4.11 Metodu neurčitých koeficientů je pochopitelně možné použít i v případě, kdy $m_0 = \dots = m_n = 1$, tedy když jsou v uzlových bodech předepsány pouze funkční hodnoty. Polynom stupně nejvýše $n - 1$, který takto najdeme, je jen jiným zápisem Lagrangeova resp. Newtonova interpolačního polynomu, jak plyne z jednoznačnosti dokázané ve větě 4.1.

Konstrukce Lagrangeova typu

Připomeňme si označení. Je dáno $n + 1$ po dvou různých uzlových bodů x_0, \dots, x_n (nemusí být uspořádané podle velikosti) a stejný počet přirozených čísel m_0, \dots, m_n . Dále je dáno $m = m_0 + \dots + m_n$ čísel označených symboly $y_i^{(j)}$, $i = 0, \dots, n$, $j = 0, \dots, m_i - 1$. Tedy pro každý uzel x_i je předepsáno právě m_i čísel. Úkolem je najít polynom $P(x)$ stupně nejvýše $m - 1$ (takový polynom má nejvýše m nenulových koeficientů), pro nějž v každém uzlovém bodu x_i platí $P^{(j)}(x_i) = y_i^{(j)}$, $j = 0, \dots, m_i - 1$. Tudíž v každém uzlovém bodu x_i je předepsáno m_i podmínek — funkční hodnota a hodnoty derivací až do řádu $m_i - 1$. Podle věty 4.9 vždy existuje právě jeden takový polynom. Konkrétní volbou čísel $y_i^{(j)}$ nyní vytvoříme speciální systém polynomů.

Předpokládejme v dalším, že uzly x_0, \dots, x_n a počty podmínek m_0, \dots, m_n jsou pevně zadané. Pro libovolnou pevně vybranou dvojici indexů i a j , $0 \leq i \leq n$, $0 \leq j \leq m_i - 1$, zvolíme $y_i^{(j)} = 1$ a pro všechny ostatní dvojice indexů r a s , $0 \leq r \leq n$, $0 \leq s \leq m_r - 1$, zvolíme $y_r^{(s)} = 0$. Pak existuje jediný polynom $L_{ij}(x)$ stupně nejvýše $m - 1$ takový, že

$$L_{ij}^{(s)}(x_r) = \begin{cases} 1 & \text{když } i = r \text{ a } j = s, & 0 \leq i \leq n, 0 \leq j \leq m_i - 1, \\ 0 & \text{v ostatních případech,} & 0 \leq r \leq n, 0 \leq s \leq m_r - 1. \end{cases} \quad (4.10)$$

Předchozí postup provedeme pro každou dvojici indexů i a j , $0 \leq i \leq n$, $0 \leq j \leq m_i - 1$. Obdržíme tak systém m polynomů $L_{ij}(x)$ stupňů nejvýše $m - 1$, které jsou plně charakterizované vztahem (4.10). Mají tedy právě jednu předepsanou funkční hodnotu resp. hodnotu derivace rovnou jedné a všechny ostatní nulové. Nazývají se *zobecněné Lagrangeovy polynomy*.

S využitím vztahu (4.10) se nyní snadno ověří, že hledaný Hermitův polynom $P(x)$ se dá vyjádřit vzorcem

$$P(x) = \sum_{i=0}^n \sum_{j=0}^{m_i-1} y_i^{(j)} L_{ij}(x). \quad (4.11)$$

Předchozí vzorec je zobecněním vztahu (4.3). Pro $m_0 = \dots = m_n = 1$ totiž bude ke každému uzlu x_i přiřazen jediný polynom $L_{i0}(x)$. Ze (4.10) je jasné, že je to právě Lagrangeův polynom $L_i(x)$ z důkazu věty 4.1.

V dalším se budeme věnovat konstrukci zobecněných Lagrangeových polynomů. Uvidíme, že je to podstatně pracnější než v případě Lagrangeových polynomů $L_i(x)$ použitých k určení Lagrangeova interpolačního polynomu ve vzorci (4.3).

Vyjdeme z m -tice speciálních polynomů $l_{ij}(x)$, $0 \leq i \leq n$, $0 \leq j \leq m_i - 1$, daných

vztahy

$$l_{ij}(x) = \frac{(x - x_i)^j}{j!} \prod_{\substack{k=0 \\ k \neq i}}^n \left(\frac{x - x_k}{x_i - x_k} \right)^{m_k}. \quad (4.12)$$

Pomocí Leibnizova vzorce pro vyšší derivace součinu dvou funkcí (viz např. [16, str. 112]) lze ověřit, že platí:

$$l_{ij}^{(j)}(x_i) = 1, \quad 0 \leq i \leq n, \quad 0 \leq j \leq m_i - 1,$$

$$l_{ij}^{(s)}(x_i) = 0, \quad 0 \leq i \leq n, \quad 0 \leq s < j \leq m_i - 1,$$

$$l_{ij}^{(s)}(x_r) = 0, \quad 0 \leq i \leq n, \quad 0 \leq j \leq m_i - 1, \quad 0 \leq r \leq n, \quad r \neq i, \quad 0 \leq s \leq m_r - 1.$$

Bohužel obecně není pravda, že $l_{ij}^{(s)}(x_i) = 0$ pro $0 \leq i \leq n, 0 \leq j < s \leq m_i - 1$, takže neplatí (4.10) a nelze položit $L_{ij}(x) = l_{ij}(x)$. Nicméně je pomocí polynomů $l_{ij}(x)$ možné rekurentně zkonstruovat zobecněné Lagrangeovy polynomy $L_{ij}(x)$. Začneme „od konce“ a splnění zmíněných neplatných podmínek dosáhneme odečtením vhodných násobků již hotových polynomů.

Položíme tedy:

$$L_{i,m_i-1}(x) = l_{i,m_i-1}(x), \quad i = 0, \dots, n, \quad (4.13)$$

a dále pak pro $j = m_i - 2, m_i - 3, \dots, 0$ položíme

$$L_{i,j}(x) = l_{i,j}(x) - \sum_{s=j+1}^{m_i-1} l_{i,j}^{(s)}(x_i) L_{i,s}(x). \quad (4.14)$$

Protože polynomy $L_{i,m_i-1}(x)$ mají stupeň $m - 1$, musí mít všechny polynomy $L_{i,j}(x)$, $j < m_i - 1$, stupeň nejvýše $m - 1$. S využitím vlastností polynomů $l_{i,j}(x)$ lze nyní ověřit, že polynomy $L_{i,j}(x)$ splňují (4.10). Našli jsme tedy skutečně jednoznačně určenou soustavu zobecněných Lagrangeových polynomů.

Příklad 4.12 Najděte Hermitův interpolační polynom v Lagrangeově tvaru $P(x)$, pro nějž platí $P(-1) = 1$, $P'(-1) = -1$, $P(0) = -2$, $P(2) = 3$ a $P'(2) = 1$.

Řešení. Jsou dány tři uzlové body $x_0 = -1$, $x_1 = 0$ a $x_2 = 2$. V prvním jsou předepsány dvě podmínky (funkční hodnota a hodnota první derivace), v druhém jedna podmínka

(funkční hodnota) a ve třetím dvě podmínky (funkční hodnota a hodnota první derivace). Při označení ze vzorce (4.11) platí $y_0^{(0)} = 1$, $y_0^{(1)} = -1$, $y_1^{(0)} = -2$, $y_2^{(0)} = 3$ a $y_2^{(1)} = 1$ a $m_0 = 2$, $m_1 = 1$ a $m_2 = 2$. Dále musíme najít pěti zobecněných Lagrangeových polynomů $L_{0,0}$, $L_{0,1}$, $L_{1,0}$, $L_{2,0}$ a $L_{2,1}$.

Nejprve sestavíme pěti pomocných polynomů l_{ij} podle vztahu (4.12):

$$\begin{aligned} l_{0,0}(x) &= \frac{(x-x_0)^0}{0!} \left(\frac{x-x_1}{x_0-x_1} \right)^{m_1} \left(\frac{x-x_2}{x_0-x_2} \right)^{m_2} = \\ &= \left(\frac{x-0}{-1-0} \right) \left(\frac{x-2}{-1-2} \right)^2 = -\frac{1}{9} x(x-2)^2, \end{aligned}$$

$$\begin{aligned} l_{0,1}(x) &= \frac{(x-x_0)^1}{1!} \left(\frac{x-x_1}{x_0-x_1} \right)^{m_1} \left(\frac{x-x_2}{x_0-x_2} \right)^{m_2} = \\ &= \frac{x-(-1)}{1} \left(\frac{x-0}{-1-0} \right) \left(\frac{x-2}{-1-2} \right)^2 = -\frac{1}{9} (x+1)x(x-2)^2, \end{aligned}$$

$$l_{1,0}(x) = \frac{(x-x_1)^0}{0!} \left(\frac{x-x_0}{x_1-x_0} \right)^{m_0} \left(\frac{x-x_2}{x_1-x_2} \right)^{m_2} =$$

$$\begin{aligned}
 &= \left(\frac{x - (-1)}{0 - (-1)} \right)^2 \left(\frac{x - 2}{0 - 2} \right)^2 = \frac{1}{4} (x + 1)^2 (x - 2)^2, \\
 l_{2,0}(x) &= \frac{(x - x_2)^0}{0!} \left(\frac{x - x_0}{x_2 - x_0} \right)^{m_0} \left(\frac{x - x_1}{x_2 - x_1} \right)^{m_1} = \\
 &= \left(\frac{x - (-1)}{2 - (-1)} \right)^2 \left(\frac{x - 0}{2 - 0} \right) = \frac{1}{18} (x + 1)^2 x, \\
 l_{2,1}(x) &= \frac{(x - x_2)^1}{1!} \left(\frac{x - x_0}{x_2 - x_0} \right)^{m_0} \left(\frac{x - x_1}{x_2 - x_1} \right)^{m_1} = \\
 &= \frac{x - 2}{1} \left(\frac{x - (-1)}{2 - (-1)} \right)^2 \left(\frac{x - 0}{2 - 0} \right) = \frac{1}{18} (x - 2)(x + 1)^2 x.
 \end{aligned}$$

Dále je ze vzorce (4.14) vidět, že budeme potřebovat hodnoty $l'_{0,0}(x_0)$ a $l'_{2,0}(x_2)$.
Připravíme si derivace a dosadíme uzly:

$$\begin{aligned}
 l'_{0,0}(x) &= -\frac{1}{9} (x - 2)^2 - \frac{2}{9} x(x - 2) & \Rightarrow & \quad l'_{0,0}(-1) = -\frac{5}{3}, \\
 l'_{2,0}(x) &= \frac{1}{9} (x + 1)x + \frac{1}{18} (x + 1)^2 & \Rightarrow & \quad l'_{2,0}(2) = \frac{7}{6}.
 \end{aligned}$$

Nyní již můžeme pomocí vzorců (4.13) a (4.14) určit zobecněné Lagrangeovy polynomy:

$$L_{0,1}(x) = l_{0,1}(x) = -\frac{1}{9}(x+1)x(x-2)^2,$$

$$L_{0,0}(x) = l_{0,0}(x) - l'_{0,0}(-1)L_{0,1}(x) = -\frac{1}{9}x(x-2)^2 - \frac{5}{27}(x+1)x(x-2)^2,$$

$$L_{1,0}(x) = l_{1,0}(x) = \frac{1}{4}(x+1)^2(x-2)^2,$$

$$L_{2,1}(x) = l_{2,1}(x) = \frac{1}{18}(x-2)(x+1)^2x,$$

$$L_{2,0}(x) = l_{2,0}(x) - l'_{2,0}(2)L_{2,1}(x) = \frac{1}{18}(x+1)^2x - \frac{7}{108}(x-2)(x+1)^2x.$$

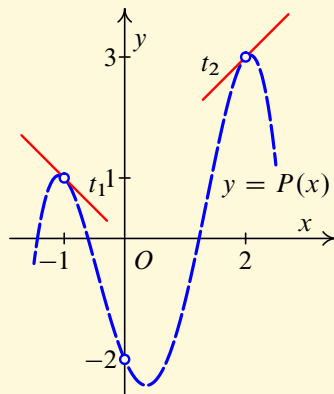
Podle vzorce (4.11) je hledaný Hermitův polynom

$$\begin{aligned} P(x) &= y_0^{(0)}L_{0,0}(x) + y_0^{(1)}L_{0,1}(x) + y_1^{(0)}L_{1,0}(x) + y_2^{(0)}L_{2,0}(x) + y_2^{(1)}L_{2,1}(x) = \\ &= \left(-\frac{1}{9}x(x-2)^2 - \frac{5}{27}(x+1)x(x-2)^2\right) - \left(-\frac{1}{9}(x+1)x(x-2)^2\right) - \end{aligned}$$

$$\begin{aligned}
 & -2\left(\frac{1}{4}(x+1)^2(x-2)^2\right) + 3\left(\frac{1}{18}(x+1)^2x - \frac{7}{108}(x-2)(x+1)^2x\right) + \\
 & + \left(\frac{1}{18}(x-2)(x+1)^2x\right) = -\frac{77}{108}x^4 + \frac{23}{18}x^3 + \frac{97}{36}x^2 - \frac{62}{27}x - 2.
 \end{aligned}$$

Jeho graf je znázorněn na obr. 4.5. Je vidět, že prochází třemi body $[-1, 1]$, $[0, -2]$ a $[2, 3]$. Navíc tečna t_1 k tomuto grafu v bodě $[-1, 1]$ má směrnici rovnou číslu -1 , protože platí rovnost $P'(-1) = -1$, tudíž její rovnice je $y - 1 = -1(x + 1)$, tj. $y = -x$, a podobně tečna t_2 v bodě $[2, 3]$ má směrnici rovnou číslu 1 , protože platí rovnost $P'(2) = 1$, tudíž její rovnice je $y - 3 = 1(x - 2)$, tj. $y = x + 1$.

Každopádně je zřejmé, že nalezení Hermitova polynomu pomocí zobecněných Lagrangeových polynomů je velmi pracné, zejména když by byly předepsané i hodnoty vyšších derivací. ▲



Obr. 4.5: $P(x) = -\frac{77}{108}x^4 + \frac{23}{18}x^3 + \frac{97}{36}x^2 - \frac{62}{27}x - 2$

Vzorce (4.14) pro polynomy $L_{i,j}(x)$ jsou rekurentní. Lze však nalézt i uzavřené vzorce a pomocí nich přepsat vzorec (4.11) pro Hermitův interpolační polynom $P(x)$.

Věta 4.13 *Za předpokladů věty 4.9 platí:*

$$L_{i,j}(x) = \left[\sum_{\ell=0}^{m_i-1} \frac{1}{\ell!} \left(\frac{\frac{1}{j!} (x-x_i)^j}{\prod_{\substack{k=0 \\ k \neq i}}^n (x-x_k)^{m_k}} \right) \Big|_{x=x_i} \right]^{(\ell)} (x-x_i)^\ell \prod_{\substack{k=0 \\ k \neq i}}^n (x-x_k)^{m_k}, \quad (4.15)$$

$$P(x) = \sum_{i=0}^n \left\{ \left[\sum_{\ell=0}^{m_i-1} \frac{1}{\ell!} \left(\frac{\sum_{j=0}^{m_i-1} \frac{y_i^{(j)}}{j!} (x-x_i)^j}{\prod_{\substack{k=0 \\ k \neq i}}^n (x-x_k)^{m_k}} \right) \Big|_{x=x_i} \right]^{(\ell)} (x-x_i)^\ell \prod_{\substack{k=0 \\ k \neq i}}^n (x-x_k)^{m_k} \right\}. \quad (4.16)$$

Pro $m_0 = \dots = m_n = 1$ přejde vzorec (4.15) ve vzorec (4.2) a vzorec (4.16) ve vzorec (4.3). Pro $n = 0$ přejde vzorec (4.16) ve vzorec (4.20), tj. Taylorův polynom.

Důkaz. Pro $i = 0, 1, \dots, n$ označme

$$Q_i(x) = \prod_{\substack{k=0 \\ k \neq i}}^n (x - x_k)^{m_k} \quad \text{a} \quad S_i(x) = \frac{\sum_{j=0}^{m_i-1} \frac{y_i^{(j)}}{j!} (x - x_i)^j}{Q_i(x)}. \quad (4.17)$$

Funkci $S_i(x)$ vyjádříme pomocí Taylorova vzorce řádu $m_i - 1$ se středem x_i . Vyjde

$$S_i(x) = \sum_{\ell=0}^{m_i-1} \frac{S_i^{(\ell)}(x_i)}{\ell!} (x - x_i)^\ell + O((x - x_i)^{m_i}).$$

Předchozí rovnost vynásobíme $Q_i(x)$ a upravíme. Protože $Q_i(x_i) \neq 0$, vyjde

$$\sum_{j=0}^{m_i-1} \frac{y_i^{(j)}}{j!} (x - x_i)^j + O((x - x_i)^{m_i}) = \sum_{\ell=0}^{m_i-1} \frac{S_i^{(\ell)}(x_i)}{\ell!} (x - x_i)^\ell Q_i(x) =: H_i(x).$$

Předchozí vztah představuje rovnost dvou polynomů, tedy výraz $O((x - x_i)^{m_i})$ je polynom mající nejméně m_i -násobný kořen x_i . Z levé strany proto plyne, že $H_i^{(j)}(x_i) = y_i^{(j)}$

pro $j = 0, \dots, m_i - 1$. Z pravé strany pak s využitím Leibnizova vzorce pro derivaci součinu plyne, že $H_i^{(j)}(x_k) = 0$ pro $j = 0, \dots, m_k - 1, k = 0, \dots, n, k \neq i$. Zřejmě st $H_i(x) \leq m - 1$, kde $m = m_0 + \dots + m_n$. Po dosazení definic $S_i(x)$ a $Q_i(x)$ do $H_i(x)$ vyjde

$$H_i(x) = \left[\sum_{\ell=0}^{m_i-1} \frac{1}{\ell!} \left(\frac{\sum_{j=0}^{m_i-1} \frac{y_i^{(j)}}{j!} (x - x_i)^j}{\prod_{\substack{k=0 \\ k \neq i}}^n (x - x_k)^{m_k}} \right) \Big|_{x=x_i} \right] (x - x_i)^\ell \prod_{\substack{k=0 \\ k \neq i}}^n (x - x_k)^{m_k}. \quad (4.18)$$

Sečtením $H_0(x) + \dots + H_n(x)$ dostaneme polynom $H(x)$ stupně nejvýše $m - 1$, který nabývá v uzlových bodech předepsané funkční hodnoty a hodnoty derivací. Vzhledem k jednoznačnosti Hermitova interpolačního polynomu (věta 4.9) musí být $H(x) = P(x)$, což dokazuje vztah (4.16).

Dále stačí zvolit pro pevné $j, 0 \leq j \leq m_i - 1$, že $y_i^{(j)} = 1$, a pro $0 \leq k \leq m_i - 1, k \neq j$, že $y_i^{(k)} = 0$. Potom $H_i(x) = L_{i,j}(x)$, čímž je dokázán vztah (4.15). \square

Příklad 4.14 Pomocí vzorce (4.16) najděte Hermitův interpolační polynom $P(x)$ pro zadání z příkladu 4.12.

Řešení. Podmínky jsou předepsány ve třech uzlech $x_0 = -1$, $x_1 = 0$ a $x_2 = 2$. Podle vzorců (4.17) tedy sestavíme funkce $S_0(x)$, $S_1(x)$ a $S_2(x)$. Protože v prvním a posledním uzlu jsou předepsány funkční hodnoty a hodnoty prvních derivací a v prostředním uzlu je předepsána pouze funkční hodnota, musíme podle vzorce (4.16) nalézt $S_0(x_0)$, $S_0'(x_0)$, $S_1(x_1)$, $S_2(x_2)$ a $S_2'(x_2)$. Jelikož $y_0^{(0)} = 1$, $y_0^{(1)} = -1$, $y_1^{(0)} = -2$, $y_2^{(0)} = 3$ a $y_2^{(1)} = 1$, postupně dostaneme:

$$S_0(x) = \frac{1 - (x + 1)}{x(x - 2)^2} = -\frac{1}{(x - 2)^2}, \quad S_0'(x) = \frac{2}{(x - 2)^3},$$

$$S_1(x) = \frac{-2}{(x + 1)^2(x - 2)^2},$$

$$S_2(x) = \frac{3 + (x - 2)}{(x + 1)^2x} = \frac{1}{x(x + 1)}, \quad S_2'(x) = \frac{-2x - 1}{x^2(x + 1)^2}.$$

Odtud vyjde:

$$S_0(-1) = -\frac{1}{9}, \quad S'_0(-1) = -\frac{2}{27}, \quad S_1(0) = -\frac{1}{2}, \quad S_2(2) = \frac{1}{6}, \quad S'_2(2) = -\frac{5}{36}.$$

Podle vzorců (4.18) a (4.16) bude

$$H_0(x) = \left(-\frac{1}{9} - \frac{2}{27}(x+1)\right)x(x-2)^2, \quad H_1(x) = -\frac{1}{2}(x+1)^2(x-2)^2,$$

$$H_2(x) = \left(\frac{1}{6} - \frac{5}{36}(x-2)\right)(x+1)^2x,$$

$$P(x) = H_0(x) + H_1(x) + H_2(x) =$$

$$= \left(-\frac{1}{9} - \frac{2}{27}(x+1)\right)x(x-2)^2 - \frac{1}{2}(x+1)^2(x-2)^2 +$$

$$+ \left(\frac{1}{6} - \frac{5}{36}(x-2)\right)(x+1)^2x = -\frac{77}{108}x^4 + \frac{23}{18}x^3 + \frac{97}{36}x^2 - \frac{62}{27}x - 2,$$

což se shoduje s výsledkem příkladu 4.12. ▲

Konstrukce Newtonova typu

Budeme předpokládat, že uzlové body jsou seřazené podle velikosti, tedy že

$$x_0 < x_1 < \cdots < x_{n-1} < x_n.$$

Nyní vytvoříme novou posloupnost, v níž každý uzel x_i zopakujeme tolikrát, kolik je pro něj předepsáno podmínek $y_i^{(j)}$. Tento počet jsme označili m_i . Vznikne tak posloupnost

$$\underbrace{x_0 = \cdots = x_0}_{m_0\text{-krát}} < \underbrace{x_1 = \cdots = x_1}_{m_1\text{-krát}} < \cdots < \underbrace{x_n = \cdots = x_n}_{m_n\text{-krát}}.$$

Nyní označíme členy této posloupnosti, jejíž délka je m , symboly t_0, t_1, \dots, t_k , kde $k = m - 1$. Tyto body nazveme *virtuální uzly*. Tedy

$$t_0 = x_0 \leq t_1 \leq \dots \leq t_k = x_n.$$

Zobecněné poměrné diference r -tého řádu, kde $r = 0, 1, \dots, k$, definujeme rekurentně takto:

Zobecněná poměrná diference nultého řádu:

$$f[t_j] = y_i^{(0)}, \quad \text{kde } t_j = x_i, \quad j = 0, \dots, k.$$

Tedy zobecněná poměrná diference nultého řádu ve virtuálním uzlu t_j je rovna předepsané funkční hodnotě v tom uzlu x_i , pro nějž platí $t_j = x_i$.

Zobecněná poměrná diference r -tého řádu, kde $r \geq 1$: Rozlišíme dva případy.

Je-li $t_j < t_{j+r}$, pak

$$f[t_j, t_{j+1}, \dots, t_{j+r}] = \frac{f[t_{j+1}, \dots, t_{j+r}] - f[t_j, \dots, t_{j+r-1}]}{t_{j+r} - t_j}, \quad j + r \leq k.$$

Je-li $t_j = t_{j+r} = x_i$, pak

$$f[t_j, t_{j+1}, \dots, t_{j+r}] = \frac{y_i^{(r)}}{r!}, \quad j + r \leq k.$$

V prvním případě je tedy zobecněná poměrná diference r -tého řádu definována pomocí zobecněných poměrných diferencí $(r - 1)$ -ního řádu, tj. obdobně jako obyčejná poměrná

diference. V druhém případě, kdy $t_j = t_{j+1} = \dots = t_{j+r}$, je zobecněná poměrná diference r -tého řádu rovna předepsané hodnotě r -té derivace v tom uzlu x_i , pro nějž platí $t_j = t_{j+1} = \dots = t_{j+r} = x_i$, dělené číslem $r!$.

Lze dokázat (viz [52, str. 56]), že Hermitův interpolační polynom je dán vzorcem

$$P(x) = f[t_0] + f[t_0, t_1](x - t_0) + f[t_0, t_1, t_2](x - t_0)(x - t_1) + \\ + \dots + f[t_0, t_1, \dots, t_k](x - t_0)(x - t_1) \dots (x - t_{k-1}). \quad (4.19)$$

Jde o obdobu (zobecnění) Newtonova interpolačního polynomu (4.5).

Výpočet zobecněných poměrných diferencí zorganizujeme obdobně jako výpočet obyčejných poměrných diferencí v tabulce 4.1, jen místo s obyčejnými uzly pracujeme s virtuálními uzly. Použití si ukážeme v následujícím příkladu.

Příklad 4.15 Najděte Hermitův interpolační polynom v Newtonově tvaru $P(x)$, pro nějž platí $P(-1) = 3$, $P'(-1) = -2$, $P''(-1) = 4$, $P(0) = 2$, $P(2) = 42$ a $P'(2) = 172$.

Řešení. S použitím předchozího označení dostaneme tři uzly $x_0 = -1$, $x_1 = 0$, $x_2 = 2$. Pro ně jsou postupně předepsány tři, jedna a dvě podmínky, takže $m_0 = 3$, $m_1 = 1$,

$m_2 = 2$ a $m = 6$. Hledaný polynom bude mít proto stupeň nejvýše 5. Dále je $y_0^{(0)} = 3$, $y_0^{(1)} = -2$, $y_0^{(2)} = 4$, $y_1^{(0)} = 2$, $y_2^{(0)} = 42$ a $y_2^{(1)} = 172$. Konečně šestice virtuálních uzlů t_0, \dots, t_5 je

$$t_0 = t_1 = t_2 = x_0 = -1, \quad t_3 = x_1 = 0, \quad t_4 = t_5 = x_2 = 2.$$

Nyní vypočteme zobecněné diference nultého až pátého řádu.

Nultý řád:

$$f[t_0] = f[t_1] = f[t_2] = y_0^{(0)} = 3,$$

$$f[t_3] = y_1^{(0)} = 2,$$

$$f[t_4] = f[t_5] = y_3^{(0)} = 42.$$

První řád:

$$f[t_0, t_1] = f[t_1, t_2] = \frac{y_0^{(1)}}{1!} = \frac{-2}{1} = -2,$$

$$f[t_2, t_3] = \frac{f[t_3] - f[t_2]}{t_3 - t_2} = \frac{2 - 3}{0 - (-1)} = -1,$$

$$f[t_3, t_4] = \frac{f[t_4] - f[t_3]}{t_4 - t_3} = \frac{42 - 2}{2 - 0} = 20,$$

$$f[t_4, t_5] = \frac{y_2^{(1)}}{1!} = \frac{172}{1} = 172.$$

Druhý řád:

$$f[t_0, t_1, t_2] = \frac{y_0^{(2)}}{2!} = \frac{4}{2} = 2,$$

$$f[t_1, t_2, t_3] = \frac{f[t_2, t_3] - f[t_1, t_2]}{t_3 - t_1} = \frac{-1 - (-2)}{0 - (-1)} = 1,$$

$$f[t_2, t_3, t_4] = \frac{f[t_3, t_4] - f[t_2, t_3]}{t_4 - t_2} = \frac{20 - (-1)}{2 - (-1)} = 7,$$

$$f[t_3, t_4, t_5] = \frac{f[t_4, t_5] - f[t_3, t_4]}{t_5 - t_3} = \frac{172 - 20}{2 - 0} = 76.$$

Třetí řád:

$$f[t_0, t_1, t_2, t_3] = \frac{f[t_1, t_2, t_3] - f[t_0, t_1, t_2]}{t_3 - t_0} = \frac{1 - 2}{0 - (-1)} = -1,$$

$$f[t_1, t_2, t_3, t_4] = \frac{f[t_2, t_3, t_4] - f[t_1, t_2, t_3]}{t_4 - t_1} = \frac{7 - 1}{2 - (-1)} = 2,$$

$$f[t_2, t_3, t_4, t_5] = \frac{f[t_3, t_4, t_5] - f[t_2, t_3, t_4]}{t_5 - t_2} = \frac{76 - 7}{2 - (-1)} = 23.$$

Čtvrtý řád:

$$f[t_0, t_1, t_2, t_3, t_4] = \frac{f[t_1, t_2, t_3, t_4] - f[t_0, t_1, t_2, t_3]}{t_4 - t_0} = \frac{2 - (-1)}{2 - (-1)} = 1,$$

$$f[t_1, t_2, t_3, t_4, t_5] = \frac{f[t_2, t_3, t_4, t_5] - f[t_1, t_2, t_3, t_4]}{t_5 - t_1} = \frac{23 - 2}{2 - (-1)} = 7.$$

Pátý řád:

$$f[t_0, t_1, t_2, t_3, t_4, t_5] = \frac{f[t_1, t_2, t_3, t_4, t_5] - f[t_0, t_1, t_2, t_3, t_4]}{t_5 - t_0} = \frac{7 - 1}{2 - (-1)} = 2.$$

Výpočty je možné přehledně uspořádat do tabulky obdobně jako v příkladu 4.6 (bez podrobného označení zobecněných diferencí, tabulka by byla příliš široká).

i	t_i	$f[t_i]$	$f[t_{i-1}, t_i]$	$f[t_{i-2}, t_{i-1}, t_i]$	$f[t_{i-3}, \dots, t_i]$	$f[t_{i-4}, \dots, t_i]$	$f[t_{i-5}, \dots, t_i]$
0	-1	3					
1	-1	3	$\frac{y_0^{(1)}}{1!} = -2$				
2	-1	3	$\frac{y_0^{(1)}}{1!} = -2$	$\frac{y_0^{(2)}}{2!} = 2$			
3	0	2	$\frac{2-3}{0-(-1)} = -1$	$\frac{-1-(-2)}{0-(-1)} = 1$	$\frac{1-2}{0-(-1)} = -1$		
4	2	42	$\frac{42-2}{2-0} = 20$	$\frac{20-(-1)}{2-(-1)} = 7$	$\frac{7-1}{2-(-1)} = 2$	$\frac{2-(-1)}{2-(-1)} = 1$	
5	2	42	$\frac{y_2^{(1)}}{1!} = 172$	$\frac{172-20}{2-0} = 76$	$\frac{76-7}{2-(-1)} = 23$	$\frac{23-2}{2-(-1)} = 7$	$\frac{7-1}{2-(-1)} = 2$

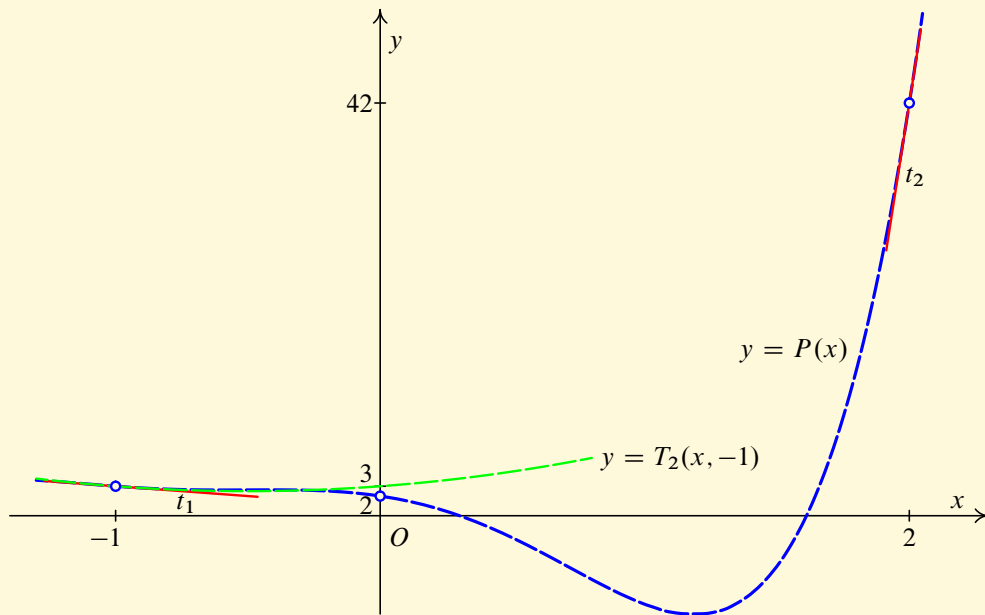
Podle vzorce (4.19) dostaneme, že

$$P(x) = f[t_0] + f[t_0, t_1](x - t_0) + f[t_0, t_1, t_2](x - t_0)(x - t_1) + \\ + f[t_0, t_1, t_2, t_3](x - t_0)(x - t_1)(x - t_2) +$$

$$\begin{aligned}
& + f[t_0, t_1, t_2, t_3, t_4](x - t_0)(x - t_1)(x - t_2)(x - t_3) + \\
& + f[t_0, t_1, t_2, t_3, t_4, t_5](x - t_0)(x - t_1)(x - t_2)(x - t_3)(x - t_4) = \\
& = 3 - 2(x + 1) + 2(x + 1)^2 - (x + 1)^3 + (x + 1)^3 x + \\
& + 2(x + 1)^3 x(x - 2) = \\
& = 3 + (x + 1) \left(-2 + (x + 1) \left(2 + (x + 1) \left(-1 + x(1 + 2(x - 2)) \right) \right) \right).
\end{aligned}$$

Zápis s postupným vytknutím je stručnější a efektivnější při dosazování za konkrétní x , srv. příklad 4.6 (dosazuje se zprava a zevnitř, jde vlastně o obdobu Hornerova schématu).

Graf $P(x)$ je znázorněn na obr. 4.6. Je vidět, že prochází třemi body $[-1, 3]$, $[0, 2]$ a $[2, 42]$. Navíc tečna t_1 k tomuto grafu v bodě $[-1, 3]$ má směrnici rovnou číslu -2 , protože platí rovnost $P'(-1) = -2$, tudíž její rovnice je $y - 3 = -2(x + 1)$, tj. $y = -2x + 1$, a podobně tečna t_2 v bodě $[2, 42]$ má směrnici rovnou číslu 172 , protože platí rovnost $P'(2) = 172$, tudíž její rovnice je $y - 42 = 172(x - 2)$, tj. $y = 172x - 302$. Vlastnost, že $P''(-1) = 4$, na grafu bezprostředně vidět není. Objeví se však v Taylorově mnohočlenu T_2 funkce $P(x)$ řádu dva se středem v bodě -1 . Je $T_2(x, -1) = P(-1) + P'(-1)(x + 1) + \frac{1}{2!}P''(-1)(x + 1)^2 = 3 - 2(x + 1) + 2(x + 1)^2$. ▲

Obr. 4.6: $P(x) = 2x^5 + 3x^4 - 4x^3 - 8x^2 - 4x + 2$

Na závěr uvedeme vzorec pro chybu při náhradě dostatečně hladké funkce Hermitovým polynomem. Platí následující zobecnění věty 4.4, viz [52, str. 57].

Věta 4.16 *Nechť uzlové body x_i , $i = 0, 1, \dots, n$, kde $n \in \mathbb{N}$, leží v intervalu $\langle a, b \rangle$, přičemž $x_0 < x_1 < \dots < x_n$. Nechť m_0, m_1, \dots, m_n jsou přirozená čísla. Označme $m = m_0 + \dots + m_n$ jejich součet.*

Předpokládejme, že funkce $f(x)$ má na intervalu $\langle a, b \rangle$ derivaci řádu m . Nechť $P(x)$ je Hermitův interpolační polynom určený podmínkami $P^{(j)}(x_i) = f^{(j)}(x_i)$, $i = 0, 1, \dots, n$, $j = 0, 1, \dots, m_i - 1$. Označme

$$f(x) = P(x) + R(x).$$

Pak pro libovolné $x \in \langle a, b \rangle$ lze chybu $R(x)$ vyjádřit ve tvaru

$$R(x) = (x - x_0)^{m_0} (x - x_1)^{m_1} \dots (x - x_n)^{m_n} \cdot \frac{f^{(m)}(\xi)}{m!},$$

kde $\xi \in \text{Int}\langle x_0, \dots, x_n, x \rangle$ je vhodné číslo. Je-li tento interval nede degenerovaný, je ξ jeho vnitřní bod (k degeneraci intervalu dojde pro $n = 0$ a $x = x_0$, pak $\xi = x_0$).

Další informace je možné najít např. v [4, 22, 43, 52].

Poznámka 4.17 V definici Hermitova polynomu se nevyklučuje, že je zadán jen jeden uzel x_0 a v něm jsou předepsány funkční hodnota a hodnoty derivací až do řádu $m - 1$. V tomto případě přejde Hermitův polynom v Taylorův¹ polynom známý ze základního kurzu diferenciálního počtu funkcí jedné proměnné (viz např. [33, str. 296]), který má při našem označení tvar

$$y_0^{(0)} + \frac{y_0^{(1)}}{1!}(x - x_0) + \frac{y_0^{(2)}}{2!}(x - x_0)^2 + \cdots + \frac{y_0^{(m-1)}}{(m-1)!}(x - x_0)^{m-1}. \quad (4.20)$$

4.2 Interpolace splajny

Jak jsme viděli v předchozím oddíle, nevýhodou interpolačního polynomu je, že graf může v koncových bodech intervalu značně oscilovat. Tato situace obvykle nastane, pokud je počet uzlových bodů velký, takže interpolační polynom je vysokého stupně. Proto se budeme zabývat funkcemi, které mají daleko lepší aproximační vlastnosti. Tyto funkce

¹**Brook Taylor** (1685–1731) (čti tejlor) — anglický matematik. Zabýval se analýzou, mechanikou a balistikou.

budou po částech tvořeny polynomy nižšího stupně, které na sebe ve společných bodech budou hladce navazovat. Takovéto funkce se nazývají *splajny*¹ a mají řadu významných aplikací. Nejčastěji se používají polynomy třetího stupně, které vedou na tzv. *kubické splajny*.

Definice 4.18 Necht' $a = x_0 < x_1 < \dots < x_n = b$ je dělení intervalu $\langle a, b \rangle$. Necht' m a ν jsou nezáporná celá čísla taková, že $0 \leq \nu \leq m + 1$. Řekneme, že funkce $S(x)$ je *splajnem stupně m s defektem ν* pro dané dělení intervalu $\langle a, b \rangle$, jestliže platí:

- 1) Na každém z intervalů $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, n - 1$, splývá funkce $S(x)$ s nějakým polynomem stupně nejvýše m .
- 2) Funkce $S(x)$ má na intervalu $\langle a, b \rangle$ spojité derivace až do řádu $m - \nu$. Přitom je-li $m = \nu$, tj. $m - \nu = 0$, požaduje se pouze spojitost funkce $S(x)$, a je-li $m + 1 = \nu$, tj. $m - \nu = -1$, požaduje se, aby funkce $S(x)$ byla po částech spojitá (nespojitost se může vyskytnout pouze v bodech x_1, \dots, x_{n-1} , musejí v nich však existovat jednostranné limity).

¹Spline je tenký úzký pružný pásek ze dřeva, gumy nebo kovu, který se v minulosti používal na kreslení velkých grafů (např. při konstrukci lodí).

Lze ukázat, že množina splajnů stupně m s defektem ν tvoří při pevně zvoleném dělení intervalu $\langle a, b \rangle$ vektorový prostor, jehož dimenze je $m + 1 + \nu(n - 1)$, viz [59, str. 17].

Kubický interpolační splajn

Nechť $a = x_0 < x_1 < \dots < x_n = b$ jsou zadané uzly a předpokládejme, že ve všech uzlech známe funkční hodnoty $y_i = f(x_i)$, $i = 0, \dots, n$, funkce $f(x)$, kterou chceme na intervalu $\langle a, b \rangle$ nahradit spljnem. Budeme hledat splajn $S(x)$ stupně tři s defektem jedna, pro nějž platí $S(x_i) = y_i$, $i = 0, \dots, n$. Takový splajn se nazývá *kubický interpolační splajn*.

Hledaný splajn $S(x)$ je funkce, která je na každém intervalu $\langle x_{i-1}, x_i \rangle$, kde $i = 1, \dots, n$, polynomem (nejvýše) třetího stupně, přičemž požadujeme, aby funkce $S(x)$, $S'(x)$ a $S''(x)$ byly na intervalu $\langle a, b \rangle$ spojité. Jednotlivé části tedy musíme napojit dostatečně hladce.

Splajn $S(x)$ bude mít následující vzorec:

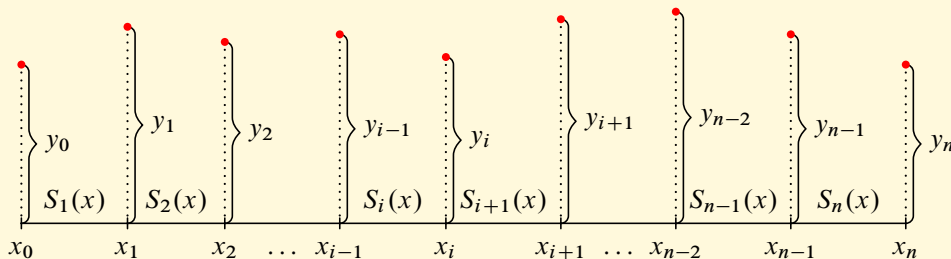
$$S(x) = \begin{cases} a_1x^3 + b_1x^2 + c_1x + d_1 = S_1(x) & \text{pro } x \in \langle x_0, x_1 \rangle, \\ a_2x^3 + b_2x^2 + c_2x + d_2 = S_2(x) & \text{pro } x \in \langle x_1, x_2 \rangle, \\ \vdots & \\ a_nx^3 + b_nx^2 + c_nx + d_n = S_n(x) & \text{pro } x \in \langle x_{n-1}, x_n \rangle. \end{cases}$$

Požadujeme, aby funkce $S(x)$ měla následující vlastnosti:

- 1) $S(x_0) = y_0, S(x_1) = y_1, \dots, S(x_n) = y_n$ (interpolační aproximace).
- 2) Funkce $S(x)$, $S'(x)$ a $S''(x)$ jsou spojité na intervalu $\langle a, b \rangle$.
- 3) Na každém intervalu $\langle x_0, x_1 \rangle, \langle x_1, x_2 \rangle, \dots, \langle x_{n-1}, x_n \rangle$ je $S(x)$ rovna polynomu nejvýše třetího stupně.

Splajn $S(x)$ je určen n -ticí nejvýše kubických polynomů $S_1(x), S_2(x), \dots, S_n(x)$. Každý z nich má čtyři koeficienty. Je tedy třeba určit $4n$ neznámých koeficientů a_1, b_1, c_1, d_1 až a_n, b_n, c_n, d_n . K jejich nalezení dostaneme z vlastností splajnu následující rovnice (viz obr. 4.7):

- a) $S_i(x_{i-1}) = y_{i-1}, S_i(x_i) = y_i, i = 1, \dots, n$ (předepsané hodnoty v uzlových bodech; tím je zajištěna i spojitost $S(x)$), tj. $2n$ rovnic.
- b) $S'_i(x_i) = S'_{i+1}(x_i), i = 1, \dots, n - 1$ (spojitost první derivace $S'(x)$), tj. $n - 1$ rovnic.
- c) $S''_i(x_i) = S''_{i+1}(x_i), i = 1, \dots, n - 1$ (spojitost druhé derivace $S''(x)$), tj. $n - 1$ rovnic.



Obr. 4.7: Kubický interpolační splajn

Snadno se zvaží, že všechny tyto rovnice jsou lineární. Dohromady tudíž máme systém $4n - 2$ lineárních rovnic o $4n$ neznámých.

Lze dokázat, že tento systém je vždy řešitelný a rovnice jsou lineárně nezávislé, tj. hodnost matice soustavy je $4n - 2$ (viz poznámka 4.19). Systém má proto nekonečně mnoho řešení a počet parametrů, na nichž řešení závisí, je roven počtu neznámých zmenšenému o hodnost matice soustavy, tj. číslu $4n - (4n - 2) = 2$. Aby bylo řešení jednoznačně určené, je tudíž třeba dodat další dvě nezávislé podmínky (tak, aby systém zůstal řešitelný). Nejvýznamnější volby (říká se jim okrajové podmínky) jsou:

- 1) $S''(x_0) = 0, S''(x_n) = 0$ — tzv. *přirozený kubický splajn*,
- 2) $S'(x_0) = f'(x_0), S'(x_n) = f'(x_n)$ — tzv. *úplný kubický splajn*
(použijeme, pokud známe hodnoty f' nebo aspoň jejich odhady v krajních bodech),
- 3) $S''(x_0) = f''(x_0), S''(x_n) = f''(x_n)$
(použijeme, pokud známe hodnoty f'' nebo aspoň jejich odhady v krajních bodech),
- 4) $S'(x_0) = S'(x_n), S''(x_0) = S''(x_n)$ — tzv. *periodický kubický splajn* s periodou $x_n - x_0$ (použijeme pro periodické funkce, musí platit $y_0 = y_n$).
- 5) Existují $S'''(x_1)$ a $S'''(x_{n-1})$ — tzv. *not a knot splajn*. V tomto případě budou mít

polynomy $S_1(x)$ a $S_2(x)$ tentýž vzorec a podobně budou mít tentýž vzorec polynomy $S_{n-1}(x)$ a $S_n(x)$. Anglický název vyjadřuje skutečnost, že v bodech x_1 a x_{n-1} tedy vlastně nejsou uzly.

K nalezení kubického splajnu $S(x)$ libovolného typu je třeba sestavit a řešit soustavu lineárních rovnic pro koeficienty jednotlivých kubických polynomů $S_i(x)$, $i = 1, \dots, n$. Každá rovnice této soustavy obsahuje maximálně šest neznámých.

Poznámka 4.19 Hledání splajnů pomocí výše zmíněné soustavy lineárních rovnic je zdlouhavé a existují daleko efektivnější postupy, jak je najít. Zmíníme se o dvou metodách, které se používají a zároveň slouží jako důkazy existence a jednoznačnosti kubických interpolačních splajnů, které splňují některou z výše uvedených dvojic okrajových podmínek.

1. První metoda je založena na tzv. kubických Hermitových splajnech. Jsou to splajny stupně tři s defektem dva. Mají tedy spojitou první derivaci. Dostanou se tak, že se kromě předepsaných podmínek $S(x_i) = y_i$, $i = 0, \dots, n$, předepíší i hodnoty prvních derivací $S'(x_i) = m_i$, $i = 0, \dots, n$. Podle věty 4.9 je na každém intervalu $\langle x_{i-1}, x_i \rangle$, $i = 1, \dots, n$, jednoznačně určen polynom $S_i(x)$ nejvýše třetího stupně takový, že $S_i(x_{i-1}) = y_{i-1}$, $S_i(x_i) = y_i$, $S'_i(x_{i-1}) = m_{i-1}$ a $S'_i(x_i) = m_i$. Tyto polynomy dávají hledaný Hermitův splajn, jenž je

uvedenými podmínkami jednoznačně určen. Z konstrukce je zřejmé, že bude mít spojitou první derivaci.

Hodnoty prvních derivací m_i v uzlových bodech však nejsou známé. Považujeme je tedy za volné parametry, které zvolíme tak, aby Hermitův splajn měl i spojitou druhou derivaci. Po přidání kterékoli z výše uvedených dvojic okrajových podmínek dostaneme pro neznámé m_i , $i = 0, \dots, n$, soustavu lineárních rovnic, jejíž matice soustavy je ryze řádkově diagonálně dominantní, takže soustava má jediné řešení (věta 3.3). Matice soustavy je třídiagonální (kromě případu periodických okrajových podmínek, kdy je navíc poslední prvek prvního řádku a první prvek posledního řádku nenulový). Pro takové soustavy existují velmi účinné modifikace GEM. Podrobněji viz [59, str. 96], pro úplné a not a knot splajny i [13, str. 49].

2. Druhá metoda vychází z toho, že druhá derivace kubického splajnu je po částech lineární spojitá funkce, která je plně určena hodnotami v uzlových bodech (grafem je lomená čára, jejíž vzorec je snadné napsat). Kromě předepsaných podmínek $S(x_i) = y_i$, $i = 0, \dots, n$, se tedy předepíše i hodnoty druhých derivací $S''(x_i) = M_i$, $i = 0, \dots, n$. Čísla M_i se nazývají momenty splajnu. Dvojí postupnou integrací na každém intervalu $\langle x_{i-1}, x_i \rangle$, kde $i = 1, \dots, n$, a dosažením uvedených podmínek se najdou jednoznačně určené nejvýše kubické polynomy $S_i(x)$ takové, že $S_i(x_{i-1}) = y_{i-1}$, $S_i(x_i) = y_i$, $S_i''(x_{i-1}) = M_{i-1}$ a $S_i''(x_i) = M_i$. Funkce sestavená z těchto polynomů však nemusí mít v uzlových bodech první derivaci, protože nemusí platit

$$S'_i(x_i) = S'_{i+1}(x_i), i = 1, \dots, n - 1.$$

Hodnoty druhých derivací M_i v uzlových bodech však nejsou známé. Považujeme je tedy za volné parametry, které zvolíme tak, aby funkce sestavená z polynomů $S_i(x)$ měla spojitou první derivaci (automaticky pak má i spojitou druhou derivaci). Po přidání kterékoli z výše uvedených dvojic okrajových podmínek dostaneme pro neznámé M_i , $i = 0, \dots, n$, soustavu lineárních rovnic, jejíž matice soustavy je ryze řádkově diagonálně dominantní, takže soustava má jediné řešení (věta 3.3). Matice soustavy má zcela analogické vlastnosti jako u první metody, založené na Hermitových splajnech. Podrobněji viz [59, str. 99], pro přirozené splajny a splajny splňující třetí typ okrajových podmínek i [41, str. 129] nebo [42, str. 45] a pro přirozené, úplné a periodické splajny [52, str. 101].

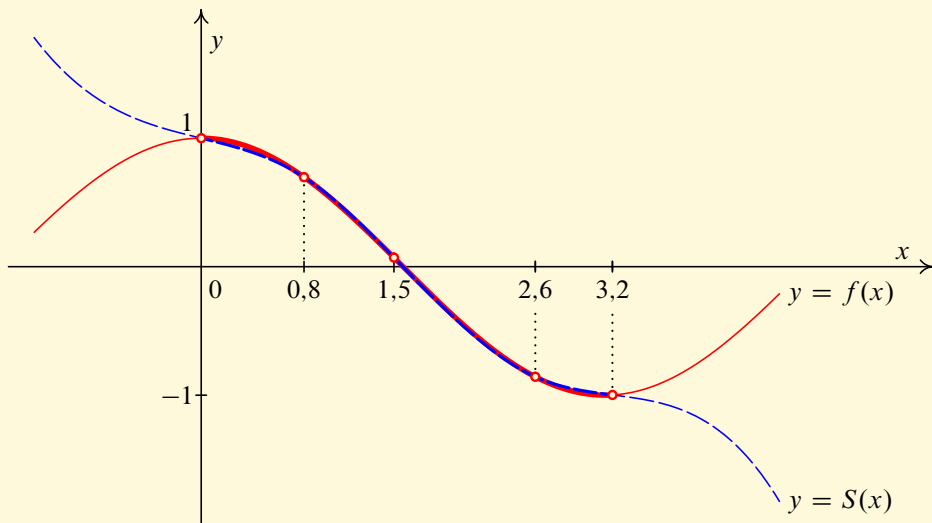
V následujících příkladech byl k nalezení koeficientů splajnu použit program počítačové algebry Maple.

Příklad 4.20 Nahraďte funkci $f(x) = \cos x$ na intervalu $\langle 0; 3,2 \rangle$ přirozeným interpolačním splajnem. Uzlové body volte $x_0 = 0$, $x_1 = 0,8$, $x_2 = 1,5$, $x_3 = 2,6$ a $x_4 = 3,2$.

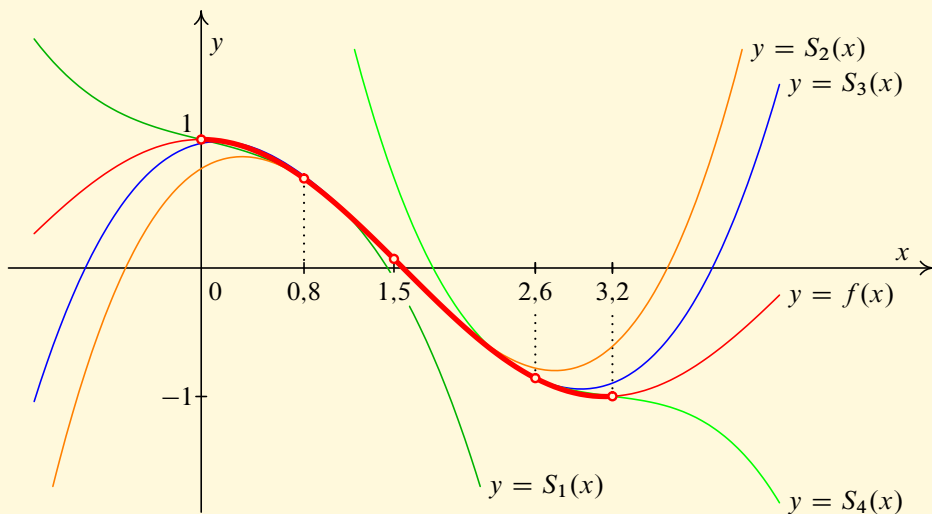
Řešení. Protože uzlových bodů je pět, bude interval $\langle 0; 3,2 \rangle$ rozdělen na čtyři podintervaly, takže musíme určit koeficienty čtyř polynomů $S_1(x)$, $S_2(x)$, $S_3(x)$ a $S_4(x)$ stupně nejvýše tři. To vede na soustavu 14 rovnic o 16 neznámých. Protože splajn má být přirozený, přibudou ještě dvě podmínky $S_1''(x_0) = 0$ a $S_4''(x_4) = 0$. Tím bude řešení určeno jednoznačně. Pomocí programu Maple obdržíme pro $S(x)$ následující vzorec (koeficienty byly zaokrouhleny na čtyři platná desetinná místa):

$$S(x) = \begin{cases} 1,0000 - 0,2433x - 0,2122x^3 = S_1(x) & \text{pro } x \leq 0,8, \\ 0,7732 + 0,6072x - 1,0631x^2 + 0,2307x^3 = S_2(x) & \text{pro } 0,8 \leq x \leq 1,5, \\ 0,9701 + 0,2134x - 0,8006x^2 + 0,1724x^3 = S_3(x) & \text{pro } 1,5 \leq x \leq 2,6, \\ 9,3136 - 9,4137x + 2,9021x^2 - 0,3023x^3 = S_4(x) & \text{pro } 2,6 \leq x. \end{cases}$$

Na obr. 4.8 je zobrazena funkce $f(x)$ a splajn $S(x)$. Je vidět, že jejich grafy téměř splývají. Na obr. 4.9 je zobrazena funkce $f(x)$ a kubické polynomy $S_1(x)$, $S_2(x)$, $S_3(x)$ a $S_4(x)$, z nichž je splajn $S(x)$ „slepen“. Je zřejmé, že polynom $S_i(x)$ je dobrou náhradou funkce $f(x)$ pouze na intervalu $\langle x_{i-1}, x_i \rangle$, $i = 1, 2, 3, 4$. ▲



Obr. 4.8: Nahrazení funkce $f(x) = \cos x$ přirozeným kubickým splajnem — grafy $f(x)$ a $S(x)$



Obr. 4.9: Nahrazení funkce $f(x) = \cos x$ přirozeným kubickým splajnem — grafy $f(x)$, $S_1(x)$, $S_2(x)$, $S_3(x)$ a $S_4(x)$

Příklad 4.21 Hodnoty neznámé funkce ve čtyřech uzlových bodech jsou dány tabulkou

x_i	-2	0	1	13/4
y_i	1	3	2	1

Najděte periodický kubický splajn pro tato data a znázorněte do jednoho obrázku jeho graf a grafy jeho prvních tří derivací.

Řešení. Protože uzlové body jsou čtyři, bude interval $\langle -2, 13/4 \rangle$ rozdělen na tři podintervaly, takže musíme určit koeficienty tří polynomů $S_1(x)$, $S_2(x)$ a $S_3(x)$ stupně nejvýše tři. To vede na soustavu 10 rovnic o 12 neznámých. Protože splajn má být periodický, přibudou ještě dvě podmínky $S'_1(x_0) = S'_3(x_3)$ a $S''_1(x_0) = S''_3(x_3)$. Tím bude řešení určeno jednoznačně. Pomocí programu Maple obdržíme pro $S(x)$ následující vzorec (koeficienty jsou racionální čísla a jsou přesné):

$$S(x) = \begin{cases} 3 - \frac{13}{63}x - \frac{9}{7}x^2 - \frac{43}{126}x^3 = S_1(x) & \text{pro } x \leq 0, \\ 3 - \frac{13}{63}x - \frac{9}{7}x^2 + \frac{31}{63}x^3 = S_2(x) & \text{pro } 0 \leq x \leq 1, \\ \frac{92}{27} - \frac{10}{7}x - \frac{4}{63}x^2 + \frac{16}{189}x^3 = S_3(x) & \text{pro } 1 \leq x. \end{cases}$$

Graf $S(x)$ je tvořen třemi oblouky kubických křivek. Pro první derivaci vyjde:

$$S'(x) = \begin{cases} -\frac{13}{63} - \frac{18}{7}x - \frac{43}{42}x^2 = S'_1(x) & \text{pro } x \leq 0, \\ -\frac{13}{63} - \frac{18}{7}x + \frac{31}{21}x^2 = S'_2(x) & \text{pro } 0 \leq x \leq 1, \\ -\frac{10}{7} - \frac{8}{63}x + \frac{16}{63}x^2 = S'_3(x) & \text{pro } 1 \leq x. \end{cases}$$

Protože derivací kubického polynomu je kvadratický polynom, je graf $S'(x)$ tvořen třemi oblouky parabol. Pro druhou derivaci vyjde:

$$S''(x) = \begin{cases} -\frac{18}{7} - \frac{43}{21}x = S''_1(x) & \text{pro } x \leq 0, \\ -\frac{18}{7} + \frac{62}{21}x = S''_2(x) & \text{pro } 0 \leq x \leq 1, \\ -\frac{8}{63} + \frac{32}{63}x = S''_3(x) & \text{pro } 1 \leq x. \end{cases}$$

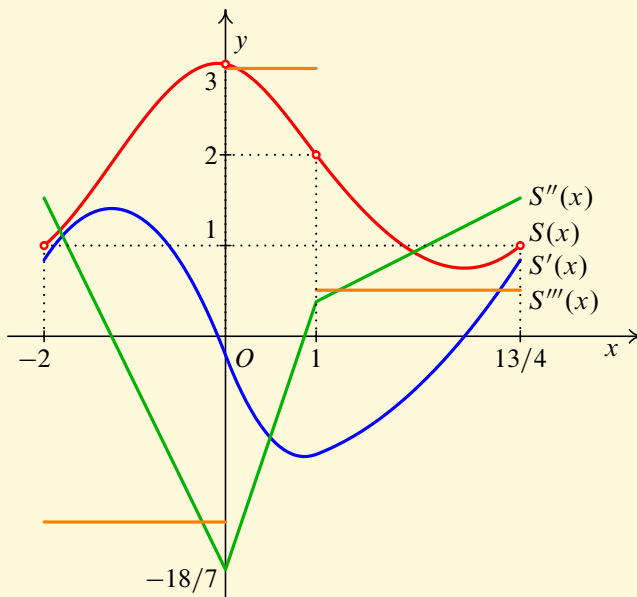
Protože derivací kvadratického polynomu je lineární polynom, je graf $S''(x)$ lomená čára tvořená třemi úsečkami.

Pro třetí derivaci vyjde:

$$S'''(x) = \begin{cases} -\frac{43}{21} = S_1'''(x) & \text{pro } x < 0, \\ \frac{62}{21} = S_2'''(x) & \text{pro } 0 < x < 1, \\ \frac{32}{63} = S_3'''(x) & \text{pro } 1 < x. \end{cases}$$

Třetí derivace neexistuje v uzlových $x_1 = 0$ a $x_2 = 1$. Protože derivací lineárního polynomu je konstanta, je graf $S'''(x)$ tvořen třemi na sebe nenavazujícími vodorovnými úsečkami (v bodech 0 a 1 má třetí derivace nespojitost prvního druhu).

Grafy splajnu $S(x)$ a jeho prvních tří derivací jsou znázorněny na obr. 4.10. ▲

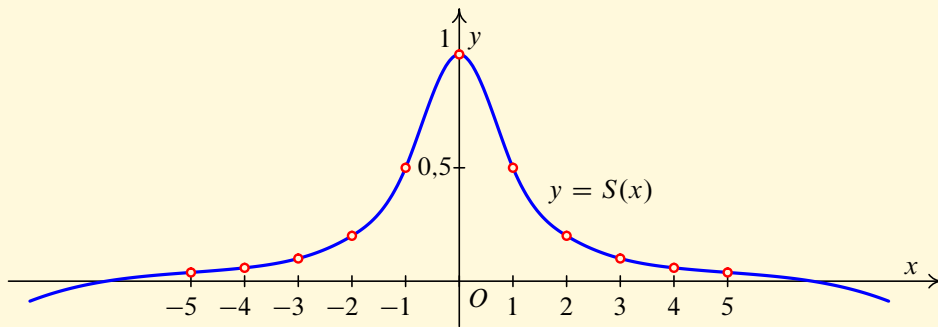
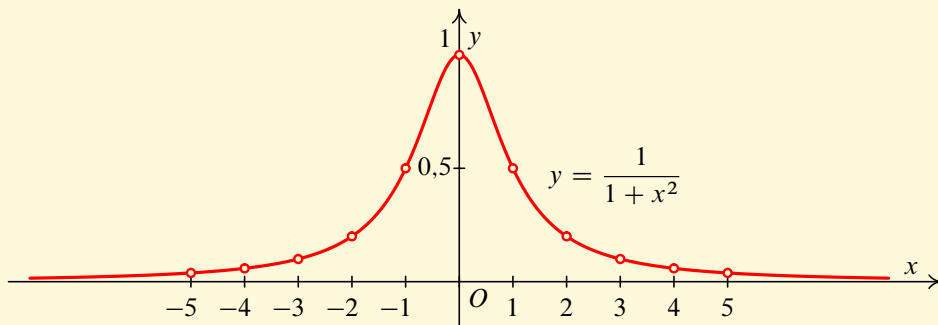
Obr. 4.10: Grafy splajnu $S(x)$ a jeho derivací $S'(x)$, $S''(x)$ a $S'''(x)$

Příklad 4.22 Nahradíte funkci $f(x) = \frac{1}{1+x^2}$ na intervalu $\langle -5, 5 \rangle$ přirozeným interpolačním splajnem. Uzlové body volte $x_i = -5, -4, \dots, 4, 5$. Porovnejte výsledek s Lagrangeovým interpolačním polynomem z příkladu 4.5.

Řešení. Protože uzlových bodů je jedenáct, bude interval $\langle -5, 5 \rangle$ rozdělen na deset podintervalů, takže musíme určit koeficienty desíti polynomů $S_1(x), S_2(x), \dots, S_{10}(x)$ stupně nejvýše tři. To vede na soustavu 38 rovnic o 40 neznámých. Protože splajn má být přirozený, přibudou ještě dvě podmínky $S_1''(-5) = 0$ a $S_{10}''(5) = 0$. Tím bude řešení určeno jednoznačně. Pomocí programu Maple obdržíme pro $S(x)$ následující vzorec (koeficienty jsou racionální čísla a jsou přesné):

$$S(x) = \begin{cases} \frac{1441}{3077} + \frac{44532}{200005}x + \frac{6561}{160004}x^2 + \frac{2187}{800020}x^3 & \text{pro } x \leq -4, \\ \frac{30029}{40001} + \frac{6684}{15385}x + \frac{15033}{160004}x^2 + \frac{5717}{800020}x^3 & \text{pro } -4 \leq x \leq -3, \\ \frac{147688}{200005} + \frac{1299}{3077}x + \frac{71889}{800020}x^2 + \frac{5353}{800020}x^3 & \text{pro } -3 \leq x \leq -2, \\ \frac{308612}{200005} + \frac{325821}{200005}x + \frac{554661}{800020}x^2 + \frac{17163}{160004}x^3 & \text{pro } -2 \leq x \leq -1, \\ 1 - \frac{748623}{800020}x^2 - \frac{348613}{800020}x^3 & \text{pro } -1 \leq x \leq 0, \\ 1 - \frac{748623}{800020}x^2 + \frac{348613}{800020}x^3 & \text{pro } 0 \leq x \leq 1 \\ \frac{308612}{200005} - \frac{325821}{200005}x + \frac{554661}{800020}x^2 - \frac{17163}{160004}x^3 & \text{pro } 1 \leq x \leq 2, \\ \frac{147688}{200005} - \frac{1299}{3077}x + \frac{71889}{800020}x^2 - \frac{5353}{800020}x^3 & \text{pro } 2 \leq x \leq 3, \\ \frac{30029}{40001} - \frac{6684}{15385}x + \frac{15033}{160004}x^2 - \frac{5717}{800020}x^3 & \text{pro } 3 \leq x \leq 4, \\ \frac{1441}{3077} - \frac{44532}{200005}x + \frac{6561}{160004}x^2 - \frac{2187}{800020}x^3 & \text{pro } 4 \leq x. \end{cases}$$

Grafy funkce $f(x)$ a splajnu $S(x)$ jsou znázorněny na obr. 4.11. Na uvažovaném intervalu téměř splývají. To je zásadní rozdíl oproti Lagrangeovu interpolačnímu polynomu (viz obr. 4.3), který je mnohem horší náhradou (v okolí krajních bodů zcela nepoužitelnou). Důvodem je příliš velký počet uzlových bodů, což v případě splajnu nevádí. ▲



Obr. 4.11: Nahrazení funkce $f(x) = \frac{1}{1+x^2}$ přirozeným kubickým splajnem

Kubické splajny jsou velmi dobrou aproximací. Jejich předností je, že se málo vlní a dávají křivky, které danými body procházejí hladce. Jejich první a druhé derivace jsou rovněž dobrou aproximací prvních a druhých derivací funkcí, které jsou dostatečně hladké (viz poznámka 4.24). Splajny mají rozsáhlé použití v numerice, počítačové grafice, zpracování signálů, statistice a dalších oblastech.

Poznámka 4.23 Přirozené splajny mají jistou důležitou extrémální vlastnost. Uvažujme systém dvojic $[x_i, y_i]$, $i = 0, \dots, n$, kde $a = x_0 < \dots < x_n = b$ je dělení intervalu $\langle a, b \rangle$. Mezi všemi funkcemi $f(x)$, které mají spojitou druhou derivaci na $\langle a, b \rangle$ (obecněji stačí, aby $f'(x)$ byla absolutně spojitá a $f''(x)$ byla lebesgueovskými integrovatelná s kvadrátem) a splňují interpolační podmínky $f(x_i) = y_i$, $i = 0, \dots, n$, nabývá integrál $\int_a^b |f''(x)|^2 dx$ nejmenší hodnoty pro přirozený kubický splajn. Protože druhá derivace $f''(x)$ souvisí s křivostí grafu funkce $f(x)$, je v jistém smyslu přirozený kubický splajn mezi funkcemi splňujícími interpolační podmínky ta funkce, která má „nejméně zakřivený graf“.

Podobně úplný resp. periodický kubický splajn minimalizuje zmíněný integrál mezi těmi výše popsány funkcemi, které navíc splňují příslušné okrajové podmínky, tj. $f'(a) = y'_0$ a $f'(b) = y'_n$, kde y'_0 a y'_n jsou daná čísla, resp. $f'(a) = f'(b)$ a $f''(a) = f''(b)$, přičemž $y_0 = y_n$. Viz [52, str. 100] nebo [59, str. 147].

Poznámka 4.24 Předpokládejme, že funkce $f(x)$ je definovaná na intervalu $\langle a, b \rangle$ a $a = x_0 < \dots < x_n = b$ je dělení tohoto intervalu, které označíme Δ . Označme $S_\Delta(x)$ kubický splajn určený hodnotami $[x_i, f(x_i)]$, $i = 0, \dots, n$ (a nějakými okrajovými podmínkami). Položme si otázku, co se stane, když budeme počet uzlových bodů zvětšovat. Na rozdíl od aproximace Lagrangeovým interpolačním polynomem, kdy kvalita aproximace mimo uzlové body se obecně výrazně zhoršuje, se ukazuje, že kubické interpolační splajny budou za velmi obecných předpokladů o funkci $f(x)$ a dělení Δ na celém intervalu $\langle a, b \rangle$ čím dál lepší náhradou.

Pro dané dělení Δ zavedeme následující označení: $\underline{h}(\Delta) = \min_{i=1, \dots, n} \{x_i - x_{i-1}\}$ (délka nejkratšího podintervalu), $\bar{h}(\Delta) = \max_{i=1, \dots, n} \{x_i - x_{i-1}\}$ (délka nejdelšího podintervalu). Číslo $\|\Delta\| = \bar{h}(\Delta)$ rovněž nazýváme *norma dělení* Δ . Pro poměr $\beta(\Delta) = \bar{h}(\Delta)/\underline{h}(\Delta)$ platí $\beta(\Delta) \geq 1$, přičemž $\beta(\Delta) = 1$ právě tehdy, když je dělení ekvidistantní. Tedy $\beta(\Delta)$ charakterizuje, jak moc se dělení Δ liší od ekvidistantního dělení.

Uvažujme nekonečnou posloupnost dělení $\{\Delta_n\}$, $n = 1, 2, \dots$ intervalu $\langle a, b \rangle$ a odpovídající posloupnost kubických interpolačních splajnů $\{S_{\Delta_n}(x)\}$, $n = 1, 2, \dots$. Řekneme, že posloupnost $\{S_{\Delta_n}(x)\}$ *konverguje stejnoměrně* na intervalu $\langle a, b \rangle$ k funkci $f(x)$, jestliže k libovolnému číslu $\varepsilon > 0$ existuje index n_0 takový, že pro $n \geq n_0$ platí $|f(x) - S_{\Delta_n}(x)| < \varepsilon$ pro libovolné $x \in \langle a, b \rangle$. Tedy grafy splajnů $S_{\Delta_n}(x)$ s dostatečně velkým indexem n leží v pásu o šířce 2ε

kolem grafu funkce $f(x)$, přičemž číslo ε můžeme volit velmi malé.

V dalším výkladu budeme potřebovat úplné splajny pro funkce, které nemusí mít derivaci. Proto příslušné okrajové podmínky nahradíme diferencemi. Je-li $\Delta: a = x_0 < \dots < x_n = b$, budeme místo podmínek $S'_\Delta(a) = f'(x_0)$ a $S'_\Delta(b) = f'(x_n)$ uvažovat podmínky

$$S'_\Delta(a) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \quad S'_\Delta(b) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}. \quad (4.21)$$

Předpokládejme nyní, že dělení Δ_n budou čím dál jemnější, tj. že bude platit $\|\Delta_n\| \rightarrow 0$ pro $n \rightarrow \infty$. Z netriviálních nerovností, které lze dokázat pro velikost rozdílu $|f(x) - S_{\Delta_n}(x)|$ v závislosti na druhu kubického splajnu a způsobu dělení intervalu $\langle a, b \rangle$, vyplývají následující výsledky:

- (1) Nechť funkce $f(x)$ je *spojitá* na $\langle a, b \rangle$, $S_{\Delta_n}(x)$ jsou buď úplné kubické splajny splňující (4.21), nebo periodické kubické splajny a existuje konstanta K taková, že $\beta(\Delta_n) < K$, $n = 1, 2, \dots$. Pak posloupnost $\{S_{\Delta_n}(x)\}$ konverguje stejnoměrně na intervalu $\langle a, b \rangle$ k funkci $f(x)$. Viz [59, str. 102].
- (2) Nechť funkce $f(x)$ má *spojitou derivaci* na $\langle a, b \rangle$ (stačí, aby $f(x)$ byla absolutně spojitá a $f'(x)$ podstatně ohraničená) a $S_{\Delta_n}(x)$ jsou buď úplné kubické splajny splňující (4.21), nebo periodické kubické splajny.

Pak posloupnost $\{S_{\Delta_n}(x)\}$ konverguje stejnoměrně na intervalu $\langle a, b \rangle$ k funkci $f(x)$. Viz [59, str. 104].

- (3) Nechť funkce $f(x)$ má *spojitou čtvrtou derivaci* na $\langle a, b \rangle$ a $S_{\Delta_n}(x)$ jsou úplné kubické splajny. Pak posloupnosti $\{S_{\Delta_n}(x)\}$, $\{S'_{\Delta_n}(x)\}$ a $\{S''_{\Delta_n}(x)\}$ konvergují stejnoměrně na intervalu $\langle a, b \rangle$ po řadě k funkcím $f(x)$, $f'(x)$ a $f''(x)$.

Existuje-li navíc konstanta K taková, že $\beta(\Delta_n) < K$, $n = 1, 2, \dots$, pak rovněž posloupnost $\{S'''_{\Delta_n}(x)\}$ konverguje stejnoměrně na intervalu $\langle a, b \rangle$ k funkci $f'''(x)$. Viz [52, str. 109].

Předchozí výsledky ukazují, že (úplné resp. periodické) kubické interpolační splajny jsou velmi dobrými aproximacemi. Navíc v případě dostatečně hladkých funkcí dobře aproximují i první a druhé derivace a pokud se použítá dělení příliš neliší od ekvidistantních, i třetí derivace.

V knize [59] lze nalézt řadu dalších nerovností, které ukazují, že i další typy kubických splajnu mají dobré aproximační vlastnosti.

4.3 Aproximace metodou nejmenších čtverců

Předpokládáme, že známe funkci $f(x)$ v $n + 1$ tabulkových bodech, ale data $y_i \approx f(x_i)$, $i = 0, \dots, n$, jsou zatížena chybami. Chceme najít aproximaci, která zachová podstatné vlastnosti funkce a odstraní „šumy“. V tomto případě použijeme metodu nejmenších čtverců.

4.3.1 Obecný případ

Hledáme aproximaci funkce $f(x)$ ve tvaru lineární kombinace

$$\varphi(x) = c_0\varphi_0(x) + \dots + c_m\varphi_m(x),$$

kde $\{\varphi_k(x)\}_{k=0}^{\infty}$ je daná posloupnost funkcí. Obvykle $m \ll n$. Nepožadujeme, aby hledaná funkce splňovala podmínky $\varphi(x_i) = y_i$, $i = 0, \dots, n$, protože se domníváme, že hodnoty y_i jsou zatížené chybami, a není důvod tyto chyby zachovat.

Chceme najít koeficienty c_0, c_1, \dots, c_m v lineární kombinaci tak, aby výraz

$$\sum_{i=0}^n (y_i - \varphi(x_i))^2 = (y_0 - \varphi(x_0))^2 + \dots + (y_n - \varphi(x_n))^2 \quad (4.22)$$

nabýval minimální hodnotu — viz. obr. 4.12. Z obrázku je zřejmé, proč se metodě říká *metoda nejmenších čtverců*.

Zavedeme následující vektory v \mathbb{R}^{n+1} :

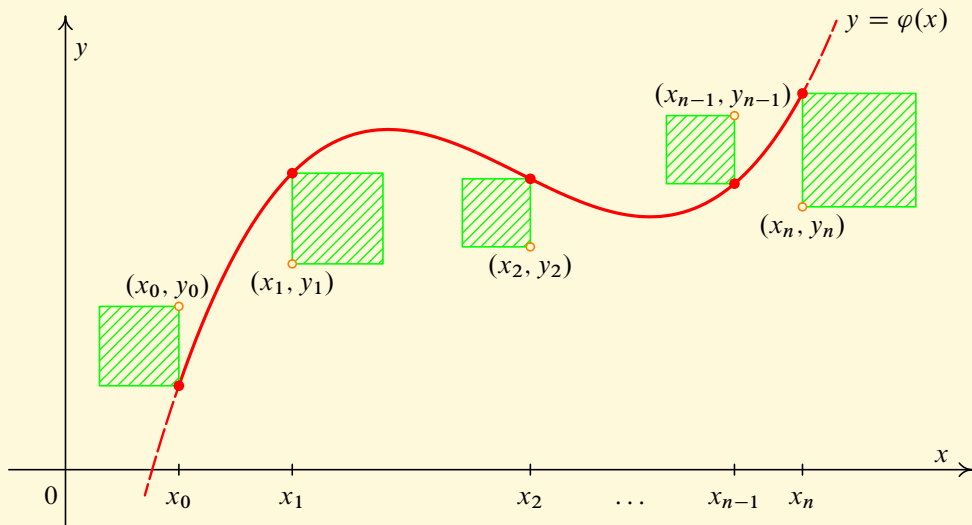
$$\widehat{\varphi}_0 = (\varphi_0(x_0), \varphi_0(x_1), \dots, \varphi_0(x_n)),$$

$$\widehat{\varphi}_1 = (\varphi_1(x_0), \varphi_1(x_1), \dots, \varphi_1(x_n)),$$

$$\vdots$$

$$\widehat{\varphi}_m = (\varphi_m(x_0), \varphi_m(x_1), \dots, \varphi_m(x_n)),$$

$$\widehat{f} = (y_0, \dots, y_n).$$



Obr. 4.12: Metoda nejmenších čtverců

Dále položíme $\widehat{\varphi} = c_0\widehat{\varphi}_0 + \dots + c_m\widehat{\varphi}_m$. Snadno se ověří, že platí

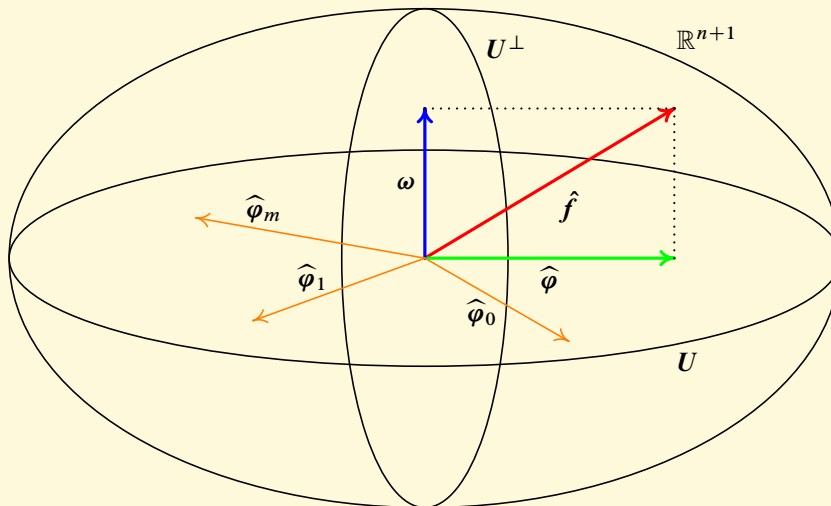
$$\|\widehat{f} - \widehat{\varphi}\|^2 = \sum_{i=0}^n (y_i - \varphi(x_i))^2,$$

kde symbol $\|\cdot\|$ značí eukleidovskou normu. Tedy úloha minimalizovat výraz (4.22) je rovnocenná minimalizaci normy $\|\widehat{f} - \widehat{\varphi}\|$. Vlastně jde o to, aby lineární kombinace $c_0\widehat{\varphi}_0 + \dots + c_m\widehat{\varphi}_m$ co nejlépe aproximovala v eukleidovské metrice vektor \widehat{f} .

Z věty 1.26 a vztahu (1.12) plyne, že $\widehat{\varphi}$ musí být pravouhlý průmět vektoru \widehat{f} na podprostor U generovaný vektory $\widehat{\varphi}_0, \dots, \widehat{\varphi}_m$ — viz obr. 4.13, kde $U = [\widehat{\varphi}_0, \dots, \widehat{\varphi}_m]$. Tedy

$$\widehat{f} = c_0\widehat{\varphi}_0 + \dots + c_m\widehat{\varphi}_m + \omega, \quad (4.23)$$

kde $\omega \in [\widehat{\varphi}_0, \dots, \widehat{\varphi}_m]^\perp = U^\perp$ (U^\perp značí ortogonální doplněk podprostoru U). Rovnici (4.23) postupně skalárně násobíme vektory $\widehat{\varphi}_0, \dots, \widehat{\varphi}_m$. Připomeňme, že standardní skalární součin vektorů $u, v \in \mathbb{R}^{n+1}$ značíme symbolem (u, v) . S využitím vlastností



Obr. 4.13: Průmět vektoru \hat{f} na podprostor $U = [\hat{\varphi}_0, \dots, \hat{\varphi}_m]$

skalárního součinu dostaneme:

$$\begin{aligned} c_0(\widehat{\varphi}_0, \widehat{\varphi}_0) + \cdots + c_m(\widehat{\varphi}_0, \widehat{\varphi}_m) &= (\widehat{\varphi}_0, \widehat{f}), \\ &\vdots \\ c_0(\widehat{\varphi}_m, \widehat{\varphi}_0) + \cdots + c_m(\widehat{\varphi}_m, \widehat{\varphi}_m) &= (\widehat{\varphi}_m, \widehat{f}), \end{aligned} \tag{4.24}$$

protože $(\widehat{\varphi}_i, \omega) = 0$ pro $i = 0, \dots, m$. Vzniklá soustava lineárních rovnic (4.24) se nazývá *normální soustava rovnic*. Matice soustavy (je symetrická) je Gramova matice vektorů $\widehat{\varphi}_0, \dots, \widehat{\varphi}_m$ a značí se $G(\widehat{\varphi}_0, \dots, \widehat{\varphi}_m)$ — viz str. 83. Tedy

$$G(\widehat{\varphi}_0, \dots, \widehat{\varphi}_m) = \begin{pmatrix} (\widehat{\varphi}_0, \widehat{\varphi}_0) & (\widehat{\varphi}_0, \widehat{\varphi}_1) & \cdots & (\widehat{\varphi}_0, \widehat{\varphi}_m) \\ (\widehat{\varphi}_1, \widehat{\varphi}_0) & (\widehat{\varphi}_1, \widehat{\varphi}_1) & \cdots & (\widehat{\varphi}_1, \widehat{\varphi}_m) \\ \dots & \dots & \dots & \dots \\ (\widehat{\varphi}_m, \widehat{\varphi}_0) & (\widehat{\varphi}_m, \widehat{\varphi}_1) & \cdots & (\widehat{\varphi}_m, \widehat{\varphi}_m) \end{pmatrix}.$$

Úloha má vždy řešení, nemusí ale být jediné, záleží na hodnotě matice soustavy $G(\widehat{\varphi}_0, \dots, \widehat{\varphi}_m)$. Jediné řešení má právě v případě, když matice soustavy je regulární, což,

jak je známo, platí právě tehdy, když vektory $\hat{\varphi}_0, \dots, \hat{\varphi}_m$ jsou lineárně nezávislé. Vektor $\hat{\varphi}$ však musí vyjít vždy stejný, protože pravouhlý průmět vektoru \hat{f} je určen jednoznačně. Je-li řešení víc (pak je jich nutně nekonečně mnoho), znamená to, že vektory $\hat{\varphi}_0, \dots, \hat{\varphi}_m$ tvoří sice systém generátorů podprostoru U , ale ne bázi, protože jsou lineárně závislé. Tedy každý vektor v U jde vyjádřit jako jejich lineární kombinace, ovšem nekonečně mnoha způsoby.

Poznámka 4.25 Označme A matici velikosti $(n + 1) \times (m + 1)$, jejíž sloupce jsou tvořeny vektory $\hat{\varphi}_0, \hat{\varphi}_1, \dots, \hat{\varphi}_m$, a B sloupcovou matici délky $n + 1$, jejíž prvky jsou y_0, y_1 až y_n . Z toho, jak je definované násobení dvou matic, je zřejmé, že normální soustavu rovnic (4.24) lze elegantně zapsat ve tvaru $A^T A C = A^T B$, kde C značí sloupec neznámých c_0, c_1, \dots, c_m . Je-li matice soustavy $A^T A = G(\hat{\varphi}_0, \dots, \hat{\varphi}_m)$ regulární, existuje k ní inverzní matice a řešení normální soustavy rovnic lze zapsat ve tvaru $C = (A^T A)^{-1} A^T B$.

Jak bylo řečeno v úvodu oddílu, cílem je odstranit z hodnot y_i „šumy“. K tomu použijeme nalezenou funkci $\varphi(x)$. Předpokládáme, že $f(x) \approx \varphi(x)$ a „opravíme“ hodnoty y_i , o nichž se domníváme, že jsou zatížené chybami, na hodnoty $\varphi(x_i)$. Říkáme, že jsme provedli *vyrovnání metodou nejmenších čtverců*. Veličina $\|\omega\| = \|\hat{f} - \hat{\varphi}\|$

(norma odchylky vektoru \hat{f} od nejlepší aproximace z podprostoru U) charakterizuje velikost opravy. Se vzrůstajícím počtem m funkcí $\varphi_i(x)$ použitých na aproximaci bude $\|\omega\|$ klesat, ale na druhé straně budou méně potlačeny chyby a hodnoty budou méně vyhlazeny. Nalezení vhodných funkcí $\varphi_i(x)$ a správného kompromisu mezi rozsahem $n + 1$ vyrovnávaných dat a číslem m je klíčové pro dosažení požadovaného cíle. V případě polynomů je doporučený postup popsán v oddílu 4.3.3.

4.3.2 Vyrovnání polynomu

Nejčastějším případem je, že funkce $\varphi(x)$ je polynom, tj.

$$\varphi(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m.$$

Označme $\varphi_k(x) = x^k$, kde $k = 0, 1, 2, \dots$. Uvažujme nyní determinant matice, jejíž řádky jsou tvořeny vektory $\hat{\varphi}_0 = (1, 1, \dots, 1)$, $\hat{\varphi}_1 = (x_0, x_1, \dots, x_n)$, \dots , $\hat{\varphi}_n = (x_0^n, x_1^n, \dots, x_n^n)$.

Z algebry je známo, že platí

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_0 & x_1 & \dots & x_n \\ \dots & \dots & \dots & \dots \\ x_0^n & x_1^n & \dots & x_n^n \end{vmatrix} = \prod_{\substack{i,j=0,\dots,n, \\ i>j}} (x_i - x_j).$$

Předchozí determinant se nazývá Vandermondův¹. Protože čísla x_i jsou navzájem různá, je tento determinant roven součinu nenulových čísel, takže je nenulový. To znamená, že vektory $\widehat{\varphi}_0, \dots, \widehat{\varphi}_n$ jsou lineárně nezávislé, a tedy pro libovolné $m \leq n$ jsou vektory $\widehat{\varphi}_0, \dots, \widehat{\varphi}_m$ rovněž nezávislé. Tudíž Gramova matice je regulární.

Pro $m = n$ dostaneme v tomto případě, tj. pro $\varphi_i(x) = x^i, i = 0, \dots, n$, Lagrangeův interpolační polynom. Vektory $\widehat{\varphi}_0, \dots, \widehat{\varphi}_n$, které jsou nezávislé, budou totiž generovat celý prostor \mathbb{R}^{n+1} , takže $U = \mathbb{R}^{n+1}$, $U^\perp = \{\mathbf{0}\}$, a tedy ω bude nulový vektor. Tudíž bude platit $\sum_{i=0}^n (y_i - \varphi(x_i))^2 = \|\hat{f} - \widehat{\varphi}\|^2 = \|\omega\|^2 = 0$, tj. $y_i = \varphi(x_i), i = 0, \dots, n$.

¹**Alexandre-Théophile Vandermonde** (1735–1796) (čti vandermond) — francouzský matematik. Zmíněný determinant se v jeho pracích nikde nevyskytuje, zřejmě jde o historický omyl v názvosloví.

Protože st $\varphi \leq n$, musí se vzhledem k jednoznačnosti interpolačního polynomu jednat o Lagrangeův polynom.

Příklad 4.26 V následující tabulce jsou dány hodnoty neznámé funkce $f(x)$:

x_i	-2	-1	0	2
y_i	3	-1	2	1

Aproximujte tuto funkci metodou nejmenších čtverců polynomem druhého stupně.

Řešení. Hledáme aproximaci funkce $f(x)$ ve tvaru lineární kombinace

$$\varphi(x) = c_0 \cdot 1 + c_1x + c_2x^2.$$

Chceme najít koeficienty c_0, c_1, c_2 tak, aby výraz

$$\sum_{i=0}^3 (y_i - \varphi(x_i))^2 = (y_0 - \varphi(x_0))^2 + (y_1 - \varphi(x_1))^2 + (y_2 - \varphi(x_2))^2 + (y_3 - \varphi(x_3))^2$$

nabýval minimální hodnoty. S použitím předchozího označení dostaneme tyto vektory v \mathbb{R}^4 :

$$\begin{aligned}\widehat{\varphi}_0 &= (1, 1, 1, 1) && \dots && \text{hodnoty funkce } \varphi_0(x) = 1 \text{ v uzlech } (-2, -1, 0, 2), \\ \widehat{\varphi}_1 &= (-2, -1, 0, 2) && \dots && \text{hodnoty funkce } \varphi_1(x) = x \text{ v uzlech } (-2, -1, 0, 2), \\ \widehat{\varphi}_2 &= (4, 1, 0, 4) && \dots && \text{hodnoty funkce } \varphi_2(x) = x^2 \text{ v uzlech } (-2, -1, 0, 2), \\ \widehat{f} &= (3, -1, 2, 1) && \dots && \text{zadané hodnoty funkce } f(x) \text{ v uzlech } (-2, -1, 0, 2).\end{aligned}$$

Využijeme poznatků z lineární algebry, a hledáme pravoúhlý průmět vektoru \widehat{f} na podprostor U generovaný vektory $\widehat{\varphi}_0, \widehat{\varphi}_1, \widehat{\varphi}_2$. Tedy

$$\widehat{f} = c_0 \widehat{\varphi}_0 + c_1 \widehat{\varphi}_1 + c_2 \widehat{\varphi}_2 + \omega, \quad (4.25)$$

kde $\omega \in [\widehat{\varphi}_0, \widehat{\varphi}_1, \widehat{\varphi}_2]^\perp = U^\perp$.

Rovnici (4.25) postupně skalárně násobíme vektory $\widehat{\varphi}_0, \widehat{\varphi}_1$ a $\widehat{\varphi}_2$. Protože platí, že

$(\hat{\varphi}_i, \omega) = 0, i = 0, 1, 2$, dostaneme:

$$c_0(\hat{\varphi}_0, \hat{\varphi}_0) + c_1(\hat{\varphi}_1, \hat{\varphi}_0) + c_2(\hat{\varphi}_2, \hat{\varphi}_0) = (\hat{\varphi}_0, \hat{f}),$$

$$c_0(\hat{\varphi}_0, \hat{\varphi}_1) + c_1(\hat{\varphi}_1, \hat{\varphi}_1) + c_2(\hat{\varphi}_2, \hat{\varphi}_1) = (\hat{\varphi}_1, \hat{f}),$$

$$c_0(\hat{\varphi}_0, \hat{\varphi}_2) + c_1(\hat{\varphi}_1, \hat{\varphi}_2) + c_2(\hat{\varphi}_2, \hat{\varphi}_2) = (\hat{\varphi}_2, \hat{f}).$$

Vypočteme potřebné skalární součiny:

$$(\hat{\varphi}_0, \hat{f}) = ((1, 1, 1, 1), (3, -1, 2, 1)) = 5,$$

$$(\hat{\varphi}_1, \hat{f}) = ((-2, -1, 0, 2), (3, -1, 2, 1)) = -3,$$

$$(\hat{\varphi}_2, \hat{f}) = ((4, 1, 0, 4), (3, -1, 2, 1)) = 15,$$

$$(\hat{\varphi}_0, \hat{\varphi}_0) = ((1, 1, 1, 1), (1, 1, 1, 1)) = 4,$$

$$(\hat{\varphi}_0, \hat{\varphi}_1) = ((1, 1, 1, 1), (-2, -1, 0, 2)) = -1,$$

$$(\hat{\varphi}_0, \hat{\varphi}_2) = ((1, 1, 1, 1), (4, 1, 0, 4)) = 9,$$

$$(\hat{\varphi}_1, \hat{\varphi}_1) = ((-2, -1, 0, 2), (-2, -1, 0, 2)) = 9,$$

$$(\widehat{\varphi}_1, \widehat{\varphi}_2) = ((-2, -1, 0, 2), (4, 1, 0, 4)) = -1,$$

$$(\widehat{\varphi}_2, \widehat{\varphi}_2) = ((4, 1, 0, 4), (4, 1, 0, 4)) = 33.$$

Po dosazení získáme soustavu rovnic

$$4c_0 - c_1 + 9c_2 = 5,$$

$$-c_0 + 9c_1 - c_2 = -3,$$

$$9c_0 - c_1 + 33c_2 = 15.$$

Řešení najdeme Gaussovou eliminační metodou. Nejprve upravíme rozšířenou matici soustavy na schodovitý tvar:

$$\begin{aligned} & \left(\begin{array}{ccc|c} 4 & -1 & 9 & 5 \\ -1 & 9 & -1 & -3 \\ 9 & -1 & 33 & 15 \end{array} \right) \sim \left(\begin{array}{ccc|c} -1 & 9 & -1 & -3 \\ 4 & -1 & 9 & 5 \\ 9 & -1 & 33 & 15 \end{array} \right) \begin{array}{l} (2) \\ (1) \end{array} \sim \\ & \sim \left(\begin{array}{ccc|c} -1 & 9 & -1 & -3 \\ 0 & 35 & 5 & -7 \\ 0 & 80 & 24 & -12 \end{array} \right) \begin{array}{l} (2) + 4(1) \\ (3) + 9(1) \end{array} \sim \left(\begin{array}{ccc|c} -1 & 9 & -1 & -3 \\ 0 & 35 & 5 & -7 \\ 0 & 20 & 6 & -3 \end{array} \right) \begin{array}{l} \\ (3) : 4 \end{array} \sim \end{aligned}$$

$$\sim \left(\begin{array}{ccc|c} -1 & 9 & -1 & -3 \\ 0 & 35 & 5 & -7 \\ 0 & 0 & 22 & 7 \end{array} \right) 7(3) - 4(2)$$

Nyní zpětným dosazením vypočteme koeficienty c_0, c_1, c_2 :

$$22c_2 = 7 \quad \Rightarrow \quad c_2 = \frac{7}{22},$$

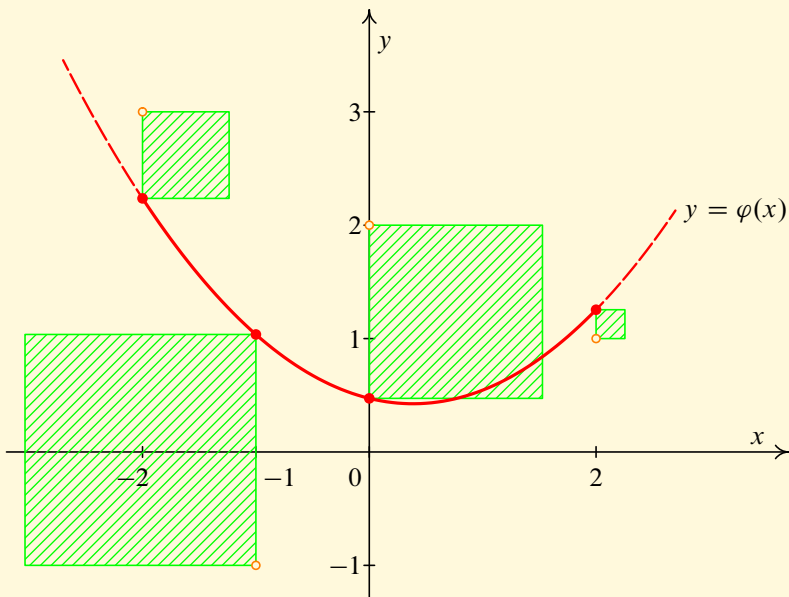
$$35c_1 + 5\frac{7}{22} = -7 \quad \Rightarrow \quad c_1 = -\frac{27}{110},$$

$$-c_0 + 9\frac{-27}{110} - \frac{7}{22} = -3 \quad \Rightarrow \quad c_0 = \frac{26}{55}.$$

Aproximační polynom druhého stupně je tvaru

$$\varphi(x) = \frac{26}{55} - \frac{27}{110}x + \frac{7}{22}x^2.$$

Jeho grafem je parabola. Výsledek je zobrazen na obr. 4.14.



Obr. 4.14: Kvadratické vyrovnání dat z příkladu 4.26 metodou nejmenších čtverců

Provedeme vyrovnání dat, tj. nahradíme hodnoty y_i funkčními hodnotami $\varphi(x_i)$. Dostaneme:

x_i	-2	-1	0	2
$\varphi(x_i)$	123/55	57/55	26/55	69/55

Odchylka je dána vektorem $\omega = \hat{f} - \hat{\varphi}$. V našem případě dostaneme

$$\|\omega\|^2 = \sum_{i=0}^3 (y_i - \varphi(x_i))^2 = \frac{392}{55} \quad \Rightarrow \quad \|\omega\| = 14 \sqrt{\frac{2}{55}} \doteq 2,67.$$

Z její velikosti a zejména z obr. 4.14 je vidět, že náhrada kvadratickým polynomem není pro daná data vhodná a že by bylo lepší použít polynom vyššího stupně (pak bychom však už dostali interpolační polynom) nebo jiný typ funkce. ▲

Poznámka 4.27 Je známo, že pro $\varphi_k(x) = x^k$, $k = 0, 1, \dots, m$, je normální soustava rovnic pro $m \geq 5$ špatně podmíněná. Proto se někdy místo báze $\hat{\varphi}_0, \hat{\varphi}_1, \dots, \hat{\varphi}_m$ používá báze tvořená diskretními ortogonálními polynomy, viz např. [29, str. 41], [43, str. 266] nebo [52, str. 235], které umožňují vyhnout se problémům se špatnou podmíněností.

Poznámka 4.28 Metoda nejmenších čtverců je dobře zdůvodněna i z pohledu statistiky. Ukazuje se, že pokud jsou chyby hodnot y_i nezávislé náhodné veličiny mající normální rozdělení s nulovou střední hodnotou a stejnou směrodatnou odchylkou $\sigma > 0$, jsou koeficienty získané metodou nejmenších čtverců maximálně věrohodnými odhady a jsou to nejlepší nestranné lineární odhady. Podrobněji viz [1, str. 188] nebo [38, str. 149].

4.3.3 Určení stupně aproximačního polynomu

Předpokládejme, že chceme vyrovnat metodou nejmenších čtverců data, o nichž se domníváme, že odpovídají přesné funkci $f(x)$, která je polynomem (nebo ji lze z hlediska praxe za polynom považovat). Pak je rozumné za aproximující funkci použít opět polynom. Důležitou otázkou ovšem je, jaký zvolit stupeň tohoto polynomu. Z předchozího výkladu vyplývá, že když zvolíme stupeň příliš malý, nebudou zachovány podstatné vlastnosti funkce $f(x)$. Zvolíme-li naopak stupeň příliš vysoký, zachovávají se i chyby vzniklé např. měřením, místo aby se data vyhladila a tyto nežádoucí „šumy“ se odstranily.

Problematikou určení správného stupně k se zabývá část statistiky nazývaná regresní analýza. Označme $\psi_k(x) = c_0 + c_1x + c_2x^2 + \dots + c_kx^k$ polynom stupně nejvýše k ,

$k = 0, 1, 2, \dots, n$, získaný metodou nejmenších čtverců z daných dat. Předpokládejme, že odchylky (chyby) hodnot y_i od přesných hodnot $f(x_i)$, $i = 0, \dots, n$, jsou nekorelované a mají všechny normální rozdělení s nulovou střední hodnotou μ a neznámou směrodatnou odchylkou $\sigma > 0$.

Určení správného k je založeno na tzv. *reziduálním rozptylu* s_k^2 definovaném vztahem

$$s_k^2 = \frac{1}{n - k} \sum_{i=0}^n (y_i - \psi_k(x_i))^2.$$

(Jmenovatel zlomku před sumou je roven rozdílu počtu uzlových bodů a počtu koeficientů polynomu $\psi_k(x)$, tj. $n + 1 - (k + 1) = n - k$.) Předpokládejme, že správný stupeň je k_0 . Lze ukázat, že pro $k < k_0$ je $\text{Es}_k^2 > \sigma^2$, kdežto pro $k \geq k_0$ je $\text{Es}_k^2 = \sigma^2$ (E značí střední hodnotu). Jde tedy o to určit, kdy se hodnoty s_k^2 přestávají v podstatě měnit. V minulosti se doporučovalo zvolit tu hodnotu k , počínaje kterou přestává reziduální rozptyl s_k^2 významně klesat, viz [42, str. 67] nebo [43, str. 265]. Taková formulace je však dost vágní.

Později se zjistilo, že místo s_k^2 je třeba vzít jiný výraz, založený sice na reziduálním rozptylu, který však ve správném k nabývá s velkou pravděpodobností minima. Používají

se následující kritéria, viz [1, str. 197], kde lze najít podrobnější výklad:

$$A_k = s_k^2 \left(1 + \frac{k+1}{\sqrt[4]{n+1}} \right), \quad (\text{Geweke a Meese, Anděl a kol.})$$

$$AIC_k = \ln s_k^2 + \frac{2(k+1)}{n+1}, \quad (\text{Akaikeovo informační kritérium})$$

$$SR_k = \ln s_k^2 + \frac{(k+1) \ln(n+1)}{n+1}, \quad (\text{Schwarz, Rissanen})$$

$$HQ_k = \ln s_k^2 + \frac{2(k+1)c \ln \ln(n+1)}{n+1}. \quad (\text{Hannan a Quinn})$$

Pro konstantu c v kritériu HQ_k musí platit $c > 1$. Někdy se volí i nepřipustná hodnota $c = 1$, obvykle se ale používá $c = 2$ nebo $c = 3$.

V praxi postupujeme tak, že stanovíme číslo $K < n$, o němž jsme z nějakých důvodů (teoretické zdůvodnění, technická omezení apod.) přesvědčení, že $k_0 \leq K$, kde k_0 je správný stupeň polynomu $f(x)$. Pak pro některé z výše uvedených kritérií určíme jeho hodnoty pro $k = 0, \dots, K$ a najdeme, pro které k nabývá nejmenší hodnotu. Toto k pak považujeme za správný stupeň.

Předchozí postup znamená řešit opakovaně normální soustavu rovnic pro určení koeficientů aproximačního polynomu, což může být numericky problematické vzhledem ke špatné podmíněnosti této soustavy pro velká k . Proto je k tomuto účelu vhodné použít diskrétní ortogonální polynomy, viz poznámka 4.27.

Příklad 4.29 Při měření funkčních hodnot polynomu zatíženém chybami byly získány následující hodnoty:

x_i	-3,8	-3,3	-2,8	-2,3	-1,8	-1,3	-0,8	-0,3
y_i	445,9	234,9	115,7	44,5	0,4982	-6,848	-6,183	-3,919
x_i	0,2	0,7	1,2	1,7	2,2	2,7	3,2	3,7
y_i	-2,067	-1,03	2,235	6,97	25,36	64,97	143,4	278,5

Odhadněte správný stupeň polynomu, je-li známo, že nemůže být větší než 8.

Řešení. Použijeme předchozí označení. Máme šestnáct měření, tedy $n = 15$. Maximální přípustný stupeň polynomu je $K = 8$. Vypočítáme reziduální rozptyly s_k^2 , $k = 0, \dots, 8$. Dále určíme hodnoty všech čtyř výše uvedených kritérií. Výpočty byly provedeny na počítači a výsledky zaokrouhlené na tři cifry jsou uvedeny v následující tabulce. V kritériu HQ_k bylo použito $c = 2$.

k	s_k^2	A_k	AIC_k	SR_k	HQ_k
0	17 300	26 000	9,88	9,93	10,0
1	17 600	35 200	10,0	10,1	10,3
2	2 320	5 800	8,12	8,27	8,52
3	2 020	6 050	8,11	8,30	8,63
4	8,42	29,5	2,76	3,00	3,41
5	9,08	36,3	2,96	3,25	3,74
6	8,73	39,3	3,04	3,38	3,95
7	9,28	46,4	3,23	3,62	4,27
8	7,64	42,0	3,16	3,59	4,33

Z tabulky je vidět, že všechna čtyři kritéria shodně určila, že správný stupeň polynomu je čtyři (pro $k = 4$ dosahují všechna kritéria minima). Také reziduální rozptyl s_k^2 přestává počínaje touto hodnotou významně klesat.

Pro získání dat ze zadání příkladu byl použit polynom

$$f(x) = 2,13x^4 - 1,57x^3 - 4,2x^2 + 6,1x - 2,4,$$

do něhož byly dosazeny hodnoty $x = -3,8; -3,3; \dots; 3,7$. K nim pak byly přičteny náhodně vygenerované chyby s normálním rozdělením o střední hodnotě $\mu = 0$ a směrodatné odchylce $\sigma = 3$. Dostali jsme tedy skutečně správnou hodnotu stupně polynomu, přestože směrodatná odchylka chyby je poměrně velká. ▲

Pojmy k zapamatování

- aproximace funkcí
- interpolační aproximace
- interpolační polynom
- Lagrangeův a Newtonův tvar interpolačního polynomu
- poměrné difference
- Newtonův interpolační polynom vpřed a vzad
- difference vpřed a vzad
- Hermitův interpolační polynom

- splajn
- kubické interpolační splajny
- přirozený, úplný, periodický a not a knot splajn
- aproximace metodou nejmenších čtverců
- pravoúhlý průmět vektoru na podprostor
- vyrovnání polynomy
- reziduální součet

Kontrolní otázky

1. Vysvětlíte pojem aproximace funkce. V jakém tvaru aproximující funkci obvykle hledáme?
2. K čemu aproximující funkce používáme?
3. Chceme aproximovat funkci, která je dána tabulkou. Mohou se v této tabulce vyskytnout např. body $[x_i, 3]$, $[x_i, 2]$?
4. Vysvětlíte, kdy použijeme k aproximaci interpolační aproximující funkci a kdy aproximaci metodou nejmenších čtverců.

5. Vysvětlete, co je interpolační polynom. Zformulujte tvrzení o jeho existenci a jednoznačnosti.
6. Jsou zadány hodnoty funkce v $n + 1$ různých uzlových bodech. Kolik různých polynomů stupně nejvýše n lze těmito body proložit? Jaká je odpověď, může-li mít polynom libovolný stupeň?
7. Jsou zadány hodnoty funkce v $n + 1$ různých uzlových bodech. Polynom nejvýše kolikátého stupně lze proložit těmito body?
8. Jaké jsou nevýhody aproximace funkcí polynomy?
9. Jak musely být rozloženy body, jestliže víte, že interpolační polynom má stupeň nula?
10. Existuje pro libovolnou čtveřici bodů $[x_i, y_i]$, kde $i = 1, 2, 3, 4$ a uzlové body x_i jsou vzájemně různé, interpolační polynom třetího stupně?
11. Uveďte vzorec Lagrangeova interpolačního polynomu pro $n + 1$ uzlových bodů.
12. Jaké vlastnosti mají speciální polynomy $L_i(x)$, které se vyskytují v Lagrangeově polynomu?
13. Uveďte vzorec Newtonova interpolačního polynomu pro $n + 1$ uzlových bodů. Jak lze určit koeficienty v tomto polynomu?
14. Napište rekurentní vztah pro výpočet poměrných diferencí $f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$ třetího řádu. Jak vypadá vzorec pro obecný n -tý řád?
15. Uveďte, kdy je vhodnější náhradou funkce Lagrangeův polynom a kdy Newtonův polynom.

16. Jaký je rozdíl mezi Lagrangeovým a Newtonovým interpolačním polynomem pro tutéž množinu bodů?
17. Napište, jak vypadá obecný tvar interpolačního polynomu nejvýše třetího stupně v Lagrangeově a Newtonově tvaru.
18. Kdy lze místo obecného Newtonova polynomu použít Newtonův polynom vpřed resp. vzad?
19. Kdy se především používají Newtonovy polynomy vpřed resp. vzad?
20. Co jsou to difference vpřed a vzad?
21. Napište vzorce pro Newtonův interpolační polynom vpřed a vzad.
22. Co je to Hermitův interpolační polynom?
23. Jaké známe metody nalezení Hermitova interpolačního polynomu?
24. Co jsou to zobecněné Lagrangeovy polynomy?
25. Uveďte vzorec Lagrangeova typu pro Hermitův interpolační polynom.
26. Co jsou to zobecněné poměrné difference?
27. Uveďte vzorec Newtonova typu pro Hermitův interpolační polynom.
28. Co jsou to splajny stupně n s defektem ν ?

29. Rozhodněte, zda je funkce $S(x)$ kubický splajn s defektem jedna s uzlovými body -2 , -1 a 1 .

$$S(x) = \begin{cases} -4x^3 - 24x^2 - 12x - 4 & \text{pro } x \leq -1, \\ 2x^3 - 6x^2 + 6x + 2 & \text{jinak.} \end{cases}$$

30. Rozhodněte, zda je funkce $S(x)$ kubický splajn s defektem jedna s uzlovými body 0 , $1/2$ a 1 .

$$S(x) = \begin{cases} 4x^3 - 9x + 12 & \text{pro } x \leq 1/2, \\ -4x^3 + 16x^2 - 15x + 12 & \text{jinak.} \end{cases}$$

31. Co je to kubický interpolační splajn?

32. Je kubický interpolační splajn určen interpolačními podmínkami jednoznačně?

33. Jaké okrajové podmínky se používají na jednoznačné určení kubického interpolačního splajnu?

34. Na jakou úlohu vede problém nalezení kubického interpolačního splajnu?

35. Uveďte příklady kubických splajnů (podle okrajových podmínek).

36. Jaké výhody má aproximace funkce kubickým splajnem oproti interpolačnímu polynomu?

37. Uveďte princip metody nejmenších čtverců.

38. Kdy je vhodné použít metodu nejmenších čtverců?

39. Co je cílem aproximace metodou nejmenších čtverců?
40. Jak je definovaná chyba, které se dopouštíme při náhradě funkce metodou nejmenších čtverců?
41. Vysvětlete geometrický význam metody nejmenších čtverců.
42. Napište, v jakém tvaru hledáme aproximující funkci u metody nejmenších čtverců.
43. Jaké typy funkcí lze použít u metody nejmenších čtverců?
44. Na jakou úlohu se převede problém nalezení koeficientů v aproximaci metodou nejmenších čtverců?
45. Co je to u metody nejmenších čtverců normální soustava rovnic?
46. Platí, že čím vyššího stupně bude polynom při aproximaci funkce metodou nejmenších čtverců, tím menší chyby se dopustíme? Zdůvodněte svou odpověď.
47. Co dostaneme, když vyrovnáme metodou nejmenších čtverců data obsahující $n + 1$ bodů polynomem stupně nejvýše n ?
48. Co je grafem aproximující funkce, provádíme-li aproximaci polynomem stupně nula, jedna, dva nebo tři?
49. Provádíme vyrovnání hodnot neznámé funkce polynomem metodou nejmenších čtverců. Jak postupujeme při určení vhodného stupně aproximujícího polynomu, víme-li, že neznámá funkce byla rovněž polynomem?

Cvičení

1. Najděte Lagrangeův interpolační polynom pro data z následující tabulky:

a)

x_i	-1	0	1	2
y_i	2	-3	1	-2

c)

x_i	-3	-1	0	2
y_i	-1	-2	1	3

e)

x_i	-1	1	2	3
y_i	-12	-2	-9	-8

g)

x_i	-7	-5	-3	-1
y_i	-2	-3	0	1

i)

x_i	-1/5	0	2/5	1
y_i	-1	1	-2	-4

k)

x_i	-2	-1	1	4
y_i	-53	-12	-2	13

b)

x_i	-1/2	0	1/2	1	2
y_i	2	-3	-2	4	3

d)

x_i	1	2	3	4	5
y_i	5	3	2	3	2

f)

x_i	-1/4	0	1/2	3/4
y_i	-1	-2	-1	1

h)

x_i	10	11	12	13
y_i	-12	-7	-9	-8

j)

x_i	-5	-4	-2	0	1
y_i	2	0	2	4	-2

l)

x_i	-1	0	1	2	3
y_i	5	3	2	3	2

m)

x_i	-2	-1	0	1	2
y_i	4	3	2	4	3

o)

x_i	3	4	5	6
y_i	-1	-2	1	3

q)

x_i	-6	-4	-2	1
y_i	1	-2	-7	-6

s)

x_i	1	2	3	4	5
y_i	-1	4	3	-3	-2

n)

x_i	-2	-1	0	1	2
y_i	2	-3	1	2	3

p)

x_i	-3/2	-1	-1/2	0	1
y_i	-1	4	3	-3	-2

r)

x_i	-1	0	1/2	2	3
y_i	2	3	4	-1	-2

t)

x_i	-1/4	0	1/4	1/2	3/4
y_i	-1	-2	-3/2	0	1/2

2. Najděte Newtonův interpolační polynom pro data z následující tabulky:

a)

x_i	0	1/2	1	3/2
y_i	1	2	-1	2

c)

x_i	-3	-1	0	2
y_i	-2	1	2	3

e)

x_i	-3	-2	0	2
y_i	-3	-1	3	-1

b)

x_i	-2	-1	0	1
y_i	1	2	-1	2

d)

x_i	-2	-1	0	1
y_i	-27	-15/2	-2	-9/2

f)

x_i	-3	-2	-1	1
y_i	45	19	7	1

g)

x_i	-1	0	1	2
y_i	1	3	-1	-5

i)

x_i	-2	-1	0	2
y_i	1	3	4	-2

k)

x_i	-7	-6	-4	-3
y_i	1	2	0	-2

m)

x_i	0	1	2	3	4
y_i	-1	2	3	-2	-1

o)

x_i	-4	-3	-1	1	2
y_i	-1	-2	0	2	3

q)

x_i	-4	-3	-1	1
y_i	-1	2	4	-2

s)

x_i	-1	0	1	2	3
y_i	-1	-2	0	2	3

h)

x_i	-7	-5	-3	-1	1
y_i	1	-2	0	-5	3

j)

x_i	-2	-1	0	1	2
y_i	-29	0	3	-2	-21

l)

x_i	-1	0	1	2	3
y_i	-1	1	-1	-2	0

n)

x_i	-3	-1	1	2	3
y_i	1	2	-1	-2	0

p)

x_i	-2	0	1	2	3
y_i	-3	-2	0	2	4

r)

x_i	4	3	1	0	5
y_i	-1	-2	1	2	2

t)

x_i	-1/2	0	1/2	1	3/2
y_i	-1	-2	0	2	3

3. Najděte Newtonův interpolační polynom vpřed a vzad pro data z následující tabulky:

a)

x_i	1	2	3	4
y_i	2,96	1,94	0,79	-0,27

b)

x_i	-20	-15	-10
y_i	78,2	93,4	61,3

c)

x_i	-2,2	-1,4	-0,6	0,2
y_i	-11,2	-28,7	-32,9	-26,4

d)

x_i	-8	-1	6
y_i	2,19	-5,27	3,64

e)

x_i	-3	0	3	6	9
y_i	2,1	5,4	7,9	6,2	5,3

f)

x_i	-5,2	-1,5	2,2	5,9
y_i	1,3	2,7	3,4	5,1

4. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 3. řádu určete přibližně $f(-1,9)$. Pomocí polynomu vzad 2. řádu určete $f(-0,1)$.

x_i	-2,2	-2,0	-1,8	-1,6	-1,4	-1,2
y_i	-0,86	-0,95	-0,47	0,18	0,70	0,96
x_i	-1,0	-0,8	-0,6	-0,4	-0,2	0,0
y_i	0,99	0,86	0,69	0,53	0,43	0,38

5. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 3. řádu určete přibližně $f(-3,7)$. Pomocí polynomu vzad 2. řádu určete $f(-2,5)$.

x_i	-4,0	-3,8	-3,6	-3,4	-3,2	-3,0	-2,8	-2,6	-2,4	-2,2
y_i	-0,64	-0,76	0,71	-0,57	-0,94	0,02	0,93	0,77	-0,12	-0,86

6. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 3. řádu určete přibližně $f(0,5)$. Pomocí polynomu vzad 3. řádu určete $f(2,2)$.

x_i	0,0	0,3	0,6	0,9	1,2	1,5
y_i	-1,20	-0,94	-0,42	0,10	0,55	0,94
x_i	1,8	2,1	2,4	2,7	3,2	
y_i	1,26	1,55	1,80	2,03	2,23	

7. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 4. řádu určete přibližně $f(-1,7)$. Pomocí polynomu vzad 2. řádu určete $f(0,3)$.

x_i	-2,0	-1,8	-1,6	-1,4	-1,2	-1,0	-0,8
y_i	-0,33	-0,35	-0,37	-0,40	-0,42	-0,44	-0,46
x_i	-0,6	-0,4	-0,2	0,0	0,2	0,4	0,6
y_i	-0,46	-0,43	-0,3	0,00	0,59	1,33	1,76

8. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 4. řádu určete přibližně $f(-2,6)$. Pomocí polynomu vzad 2. řádu určete $f(-0,1)$.

x_i	-3,0	-2,7	-2,4	-2,1	-1,8	-1,5	-1,2	-0,9	-0,6	-0,3	0,0
y_i	-0,71	-0,66	-0,61	-0,53	-0,44	-0,31	-0,14	0,08	0,37	0,67	0,88

9. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 2. řádu určete přibližně $f(-2,7)$. Pomocí polynomu vzad 4. řádu určete $f(-1,3)$.

x_i	-2,8	-2,6	-2,4	-2,2	-2,0	-1,8
y_i	5,56	1,91	-2,80	-4,40	-2,75	0,01
x_i	-1,6	-1,4	-1,2	-1,0	-0,8	
y_i	2,02	2,68	2,34	1,57	0,82	

10. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 4. řádu určete přibližně $f(-1,35)$. Pomocí polynomu vzad 4. řádu určete $f(-0,45)$.

x_i	-1,5	-1,4	-1,3	-1,2	-1,1	-1,0
y_i	0,93	1,14	1,29	1,39	1,44	1,45
x_i	-0,9	-0,8	-0,7	-0,6	-0,5	-0,4
y_i	1,44	1,40	1,34	1,28	1,21	1,14

11. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 4. řádu určete přibližně $f(-0,85)$. Pomocí polynomu vzad 3. řádu určete $f(0,05)$.

x_i	-1,0	-0,9	-0,8	-0,7	-0,6	-0,5
y_i	-0,60	-0,36	-0,16	0,01	0,16	0,27
x_i	-0,4	-0,3	-0,2	-0,1	0,0	0,1
y_i	0,36	0,43	0,47	0,49	0,50	0,49

12. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 3. řádu určete přibližně $f(-4,3)$. Pomocí polynomu vzad 2. řádu určete $f(-1,2)$.

x_i	-5,0	-4,6	-4,2	-3,8	-3,4	-3,0
y_i	0,55	-0,21	-0,89	-1,39	-1,63	-1,59
x_i	-2,6	-2,2	-1,8	-1,4	-1,0	-0,6
y_i	-1,31	-0,84	-0,30	0,21	0,59	0,79

13. V následující tabulce jsou dány hodnoty funkce f . Pomocí Newtonova polynomu vpřed 3. řádu určete přibližně $f(-2,6)$. Pomocí polynomu vzad 4. řádu určete $f(-0,6)$.

x_i	-3,0	-2,75	-2,5	-2,25	-2,0	-1,75	-1,5
y_i	-2,85	-2,54	-1,96	-1,26	-0,63	-0,18	0,04
x_i	-1,25	-1,0	-0,75	-0,5	-0,25	0,0	
y_i	0,04	-0,08	-0,21	-0,26	-0,19	0,00	

14. Metodou neurčitých koeficientů najděte Hermitův interpolační polynom P splňující následující podmínky:
- $P(-1) = 1, P(1) = 0, P'(1) = 2, P(3) = -1,$
 - $P(-2) = 0, P'(-2) = 1, P(-1) = 2, P(0) = -2, P(1) = 0, P'(1) = -1,$
 - $P(-1) = 2, P'(-1) = -1, P(1) = -3, P'(1) = 0,$
 - $P(0) = 1, P'(0) = 1, P''(0) = 1, P(1) = -1, P(2) = 2,$
 - $P(0) = 1, P'(0) = -1, P(1) = 0, P'(1) = 1, P''(1) = 2,$
 - $P(0) = 3, P'(0) = 2, P''(0) = 1, P(1) = -2, P'(1) = 0, P''(1) = 3.$

15. Pomocí zobecněných Lagrangeových polynomů (vypište je) najděte Hermitův interpolační polynom P splňující následující podmínky:

- a) $P(0) = 2, P'(0) = -1, P(1) = 3, P'(1) = -2,$
- b) $P(-1) = 1, P(1) = 2, P'(1) = -2,$
- c) $P(0) = -1, P'(0) = 1, P''(0) = 2, P(2) = -3,$
- d) $P(-1) = 2, P(0) = 3, P'(0) = -1, P(1) = -3,$
- e) $P(-1) = 2, P'(-1) = 1, P(1) = -2, P'(1) = 2,$
- f) $P(-1) = 4, P'(-1) = -2, P''(-1) = -4, P(0) = 2, P'(0) = 3.$

Řešte všechna zadání rovněž pomocí vzorce (4.16).

16. Najděte Newtonův tvar Hermitova interpolačního polynomu P splňujícího následující podmínky:

- a) $P(-2) = -3, P'(-2) = 1, P(1) = 18, P(2) = 97,$
- b) $P(-1) = -2, P'(-1) = 3, P(0) = -3, P'(0) = -4, P(3) = 426,$
- c) $P(0) = 2, P'(0) = -3, P''(0) = 8, P(1) = 1, P(2) = -28, P'(2) = -63,$
- d) $P(-3) = 2, P(-1) = -2, P'(-1) = -4, P(2) = 112, P'(2) = 197,$
- e) $P(0) = 4, P(1) = 6, P'(1) = 3, P''(1) = 0, P(2) = 4,$
- f) $P(1) = 3, P'(1) = 2, P''(1) = 2, P'''(1) = -6, P(2) = 6, P'(2) = 7.$

17. Data z tabulky aproximujte polynomem 1. stupně metodou nejmenších čtverců. Vypočítejte eukleidovskou normu vektoru odchylek (odmocninu ze součtu kvadrátů odchylek).

a)

x_i	-2	-1	1	2
y_i	2	1	2	3

c)

x_i	-3	-2	0	1
y_i	1	2	3	4

e)

x_i	-2	-1	1	2
y_i	3	1	0	1

g)

x_i	-2	-1	1	2
y_i	1	0	1	2

i)

x_i	-1	0	1	2
y_i	1	2	1	-2

b)

x_i	-3	-2	-1	0	1
y_i	1	2	2	3	4

d)

x_i	-2	-1	0	1	2
y_i	1	2	3	4	3

f)

x_i	-1	0	1	2
y_i	3	2	1	2

h)

x_i	-1	0	1	2
y_i	0	0	1	2

j)

x_i	-2	-1	0	1	2
y_i	0	2	1	-1	-2

18. Data z tabulky aproximujte polynomelem 2. stupně metodou nejmenších čtverců. Vypočtete eukleidovskou normu vektoru odchylek (odmocninu ze součtu kvadrátů odchylek).

a)

x_i	-2	-1	0	2
y_i	-1	-2	-1	7

c)

x_i	-2	-1	0	1	2
y_i	-7	-2	1	2	1

e)

x_i	-2	-1	0	2
y_i	-7	-2	1	1

g)

x_i	-2	-1	0	2
y_i	-9	-4	-1	-1

i)

x_i	-2	-1	0	2
y_i	1	0	1	9

b)

x_i	-2	-1	0	1	2
y_i	-7	0	3	2	-3

d)

x_i	-3	-2	-1	0	2
y_i	13	5	-1	-5	-7

f)

x_i	-2	-1	0	1	3
y_i	5	1	-1	-1	5

h)

x_i	-1	0	2	3	4	5
y_i	3/2	-1	-3	-5/2	-1	3/2

j)

x_i	-1	0	2	3	4	5
y_i	4	2	1	2	4	7

Řešení

1. a) $-\frac{8}{3}x^3 + \frac{9}{2}x^2 + \frac{13}{6}x - 3,$
 b) $-\frac{16}{5}x^4 + \frac{28}{15}x^3 + \frac{64}{5}x^2 - \frac{67}{15}x - 3,$
 c) $-\frac{11}{30}x^3 - \frac{3}{10}x^2 + \frac{46}{15}x + 1,$
 d) $-\frac{5}{24}x^4 + \frac{9}{4}x^3 - \frac{187}{24}x^2 + \frac{35}{4}x + 2,$
 e) $2x^3 - 8x^2 + 3x + 1,$
 f) $8x^2 - 2x - 2,$
 g) $-\frac{1}{8}x^3 - \frac{11}{8}x^2 - \frac{27}{8}x - \frac{9}{8},$
 h) $\frac{5}{3}x^3 - \frac{117}{2}x^2 + \frac{4091}{6}x - 2647,$
 i) $\frac{250}{9}x^3 - \frac{625}{18}x^2 + \frac{35}{18}x + 1,$
 j) $-\frac{2}{45}x^4 - \frac{31}{45}x^3 - \frac{26}{9}x^2 - \frac{107}{45}x + 4,$
 k) $2x^3 - 8x^2 + 3x + 1,$
 l) $-\frac{5}{24}x^4 + \frac{7}{12}x^3 + \frac{17}{24}x^2 - \frac{25}{12}x + 3,$
 m) $-\frac{3}{8}x^4 - \frac{1}{4}x^3 + \frac{15}{8}x^2 + \frac{3}{4}x + 2,$
 n) $\frac{5}{8}x^4 - \frac{3}{4}x^3 - \frac{17}{8}x^2 + \frac{13}{4}x + 1,$
 o) $-\frac{5}{6}x^3 + 12x^2 - \frac{325}{6}x + 76,$
 p) $\frac{16}{5}x^4 + \frac{164}{15}x^3 + \frac{4}{5}x^2 - \frac{209}{15}x - 3,$
 q) $\frac{7}{60}x^3 + \frac{23}{20}x^2 + \frac{17}{15}x - \frac{42}{5},$
 r) $\frac{26}{45}x^4 - \frac{89}{45}x^3 - \frac{34}{45}x^2 + \frac{14}{5}x + 3,$
 s) $\frac{11}{24}x^4 - \frac{53}{12}x^3 + \frac{289}{24}x^2 - \frac{85}{12}x - 2,$
 t) $-16x^4 + \frac{8}{3}x^3 + 13x^2 - \frac{7}{6}x - 2.$
2. a) $1 + x\left(2 + \left(x - \frac{1}{2}\right)\left(-8 + \frac{40}{3}(x - 1)\right)\right),$
 b) $1 + (x + 2)\left(1 + (x + 1)\left(-2 + \frac{5}{3}x\right)\right),$
 c) $-2 + (x + 3)\left(\frac{3}{2} - \frac{1}{6}(x + 1)\right),$

$$d) -27 + (x + 2)\left(\frac{39}{2} + (x + 1)(-7 + x)\right),$$

$$e) -3 + (x + 3)\left(2 - \frac{1}{5}(x + 2)x\right),$$

$$f) 45 + (x + 3)\left(-26 + (x + 2)(7 - (x + 1))\right),$$

$$g) 1 + (x + 1)\left(2 + x(-3 + (x - 1))\right),$$

$$h) 1 + (x + 7)\left(-\frac{3}{2} + (x + 5)\left(\frac{5}{8} + (x + 3)\left(-\frac{1}{4} + \frac{1}{12}(x + 1)\right)\right)\right),$$

$$i) 1 + (x + 2)\left(2 + (x + 1)\left(-\frac{1}{2} - \frac{5}{24}x\right)\right),$$

$$j) -29 + (x + 2)\left(29 + (x + 1)\left(-13 + x(3 - (x - 1))\right)\right),$$

$$k) 1 + (x + 7)\left(1 + (x + 6)\left(-\frac{2}{3} + \frac{1}{12}(x + 4)\right)\right),$$

$$l) -1 + (x + 1)\left(2 + x\left(-2 + (x - 1)\left(\frac{5}{6} - \frac{1}{8}(x - 2)\right)\right)\right),$$

$$m) -1 + x\left(3 + (x - 1)\left(-1 + (x - 2)\left(-\frac{2}{3} + \frac{2}{3}(x - 3)\right)\right)\right),$$

$$\text{n) } 1 + (x + 3) \left(\frac{1}{2} + (x + 1) \left(-\frac{1}{2} + (x - 1) \left(\frac{2}{15} + \frac{1}{30}(x - 2) \right) \right) \right),$$

$$\text{o) } -1 + (x + 4) \left(-1 + (x + 3) \left(\frac{2}{3} + (x + 1) \left(-\frac{2}{15} + \frac{1}{45}(x - 1) \right) \right) \right),$$

$$\text{p) } -3 + (x + 2) \left(\frac{1}{2} + x \left(\frac{1}{2} + (x - 1) \left(-\frac{1}{8} + \frac{1}{40}(x - 2) \right) \right) \right),$$

$$\text{q) } -1 + (x + 4) \left(3 + (x + 3) \left(-\frac{2}{3} - \frac{1}{15}(x + 1) \right) \right),$$

$$\text{r) } -1 + (x - 4) \left(1 + (x - 3) \left(\frac{5}{6} + (x - 1) \left(\frac{1}{4} - \frac{1}{24}x \right) \right) \right),$$

$$\text{s) } -1 + (x + 1) \left(-1 + x \left(\frac{3}{2} + (x - 1) \left(-\frac{1}{2} + \frac{1}{12}(x - 2) \right) \right) \right),$$

$$\text{t) } -1 + \left(x + \frac{1}{2} \right) \left(-2 + x \left(6 + \left(x - \frac{1}{2} \right) \left(-4 + \frac{4}{3}(x - 1) \right) \right) \right).$$

$$3. \text{ a) } N^+(t) = 2,96 - \frac{1,02}{1!}t - \frac{0,13}{2!}t(t - 1) + \frac{0,22}{3!}t(t - 1)(t - 2), \quad t = x - 1,$$

$$N^-(s) = -0,27 - \frac{1,06}{1!}s + \frac{0,09}{2!}s(s + 1) + \frac{0,22}{3!}s(s + 1)(s + 2), \quad s = x - 4,$$

$$\text{b) } N^+(t) = 78,2 + \frac{15,2}{1!}t - \frac{47,3}{2!}t(t-1), \quad t = \frac{x+20}{5},$$

$$N^-(s) = 61,3 - \frac{32,1}{1!}s - \frac{47,3}{2!}s(s+1), \quad s = \frac{x+10}{5},$$

$$\text{c) } N^+(t) = -11,2 - \frac{17,5}{1!}t + \frac{13,3}{2!}t(t-1) - \frac{2,6}{3!}t(t-1)(t-2), \quad t = \frac{x+2,2}{0,8},$$

$$N^-(s) = -26,4 + \frac{6,5}{1!}s + \frac{10,7}{2!}s(s+1) - \frac{2,6}{3!}s(s+1)(s+2), \quad s = \frac{x-0,2}{0,8},$$

$$\text{d) } N^+(t) = 2,19 - \frac{7,46}{1!}t + \frac{16,37}{2!}t(t-1), \quad t = \frac{x+8}{7},$$

$$N^-(s) = 3,64 + \frac{8,91}{1!}s + \frac{16,37}{2!}s(s+1), \quad s = \frac{x-6}{7},$$

$$\text{e) } N^+(t) = 2,1 + \frac{3,3}{1!}t - \frac{0,8}{2!}t(t-1) - \frac{3,4}{3!}t(t-1)(t-2) + \\ + \frac{8,4}{4!}t(t-1)(t-2)(t-3), \quad t = \frac{x+3}{3},$$

$$N^-(s) = 5,3 - \frac{0,9}{1!}s + \frac{0,8}{2!}s(s+1) + \frac{5,0}{3!}s(s+1)(s+2) + \\ + \frac{8,4}{4!}s(s+1)(s+2)(s+3), \quad s = \frac{x-9}{3},$$

$$\text{f) } N^+(t) = 1,3 + \frac{1,4}{1!}t - \frac{0,7}{2!}t(t-1) + \frac{1,7}{3!}t(t-1)(t-2), \quad t = \frac{x+5,2}{3,7},$$

$$N^-(s) = 5,1 + \frac{1,7}{1!}s + \frac{1}{2!}s(s+1) + \frac{1,7}{3!}s(s+1)(s+2), \quad s = \frac{x-5,9}{3,7}.$$

4. $f(-1,9) = -0,756$, $f(-0,1) = 0,399$.
5. $f(-3,7) = 0,902$, $f(-2,5) = 0,306$.
6. $f(0,5) = -0,605$, $f(2,2) = 1,63$.
7. $f(-1,7) = -0,359$, $f(0,3) = 0,999$.
8. $f(-2,6) = -0,645$, $f(-0,1) = 0,820$.
9. $f(-2,7) = 3,87$, $f(-1,3) = 2,62$.
10. $f(-1,35) = 1,22$, $f(-0,45) = 1,17$.
11. $f(-0,85) = -0,256$, $f(0,05) = 0,498$.
12. $f(-4,3) = -0,733$, $f(-1,2) = 0,423$.
13. $f(-2,6) = -2,21$, $f(-0,6) = -0,253$.
14. a) $-\frac{5}{8}x^3 + \frac{15}{8}x^2 + \frac{1}{8}x - \frac{11}{8}$, b) $-x^5 - \frac{8}{3}x^4 + \frac{5}{3}x^3 + \frac{17}{3}x^2 - \frac{5}{3}x - 2$,
c) $x^3 + \frac{1}{4}x^2 - \frac{7}{2}x - \frac{3}{4}$, d) $\frac{25}{8}x^4 - \frac{53}{8}x^3 + \frac{1}{2}x^2 + x + 1$,
e) $-3x^4 + 8x^3 - 5x^2 - x + 1$, f) $-35x^5 + \frac{179}{2}x^4 - 62x^3 + \frac{1}{2}x^2 + 2x + 3$.

15. a) $L_{0,1}(x) = x(x-1)^2$, $L_{0,0}(x) = (x-1)^2 + 2L_{0,1}(x)$, $L_{1,1}(x) = (x-1)x^2$,
 $L_{1,0}(x) = x^2 - 2L_{1,1}(x)$, $P(x) = -5x^3 + 7x^2 - x + 2$,
- b) $L_{0,0}(x) = \frac{1}{4}(x-1)^2$, $L_{1,1}(x) = \frac{1}{2}(x-1)(x+1)$,
 $L_{1,0}(x) = \frac{1}{2}(x+1) - \frac{1}{2}L_{1,1}(x)$. $P(x) = -\frac{5}{4}x^2 + \frac{1}{2}x + \frac{11}{4}$,
- c) $L_{0,2}(x) = -\frac{1}{4}x^2(x-2)$, $L_{0,1}(x) = -\frac{1}{2}x(x-2) + L_{0,2}(x)$,
 $L_{0,0}(x) = -\frac{1}{2}(x-2) + \frac{1}{2}L_{0,1}(x) + 0L_{0,2}(x)$, $L_{1,0}(x) = \frac{1}{8}x^3$,
 $P(x) = -x^3 + x^2 + x - 1$,
- d) $L_{0,0}(x) = -\frac{1}{2}x^2(x-1)$, $L_{1,1}(x) = -x(x+1)(x-1)$,
 $L_{1,0}(x) = -(x+1)(x-1) + 0L_{1,1}(x)$, $L_{2,0}(x) = \frac{1}{2}(x+1)x^2$,
 $P(x) = -\frac{3}{2}x^3 - \frac{7}{2}x^2 - x + 3$,
- e) $L_{0,1}(x) = \frac{1}{4}(x+1)(x-1)^2$, $L_{0,0}(x) = \frac{1}{4}(x-1)^2 + L_{0,1}(x)$,
 $L_{1,1}(x) = \frac{1}{4}(x-1)(x+1)^2$, $L_{1,0}(x) = \frac{1}{4}(x+1)^2 - L_{1,1}(x)$,
 $P(x) = \frac{7}{4}x^3 + \frac{1}{4}x^2 - \frac{15}{4}x - \frac{1}{4}$,
- f) $L_{0,2}(x) = \frac{1}{2}(x+1)^2x^2$, $L_{0,1}(x) = (x+1)x^2 + 4L_{0,2}(x)$,
 $L_{0,0}(x) = x^2 + 2L_{0,1}(x) - 2L_{0,2}(x)$, $L_{1,1}(x) = x(x+1)^3$,
 $L_{1,0}(x) = (x+1)^3 - 3L_{1,1}(x)$, $P(x) = 3x^4 + 11x^3 + 13x^2 + 3x + 2$.

16. a) $P(x) = -3 + (x + 2)\left(1 + (x + 2)(2 + 4(x - 1))\right),$
 b) $P(x) = -2 + (x + 1)\left(3 + (x + 1)(-4 + x(1 + 3x))\right),$
 c) $P(x) = 2 + x\left(-3 + x\left(4 + x\left(-2 + (x - 1)(-3 + (x - 2))\right)\right)\right),$
 d) $P(x) = 2 + (x + 3)\left(-2 + (x + 1)\left(-1 + (x + 1)(3 + 2(x - 2))\right)\right),$
 e) $P(x) = 4 + x\left(2 + (x - 1)\left(1 + (x - 1)(-1 - 2(x - 1))\right)\right),$
 f) $P(x) = 3 + (x - 1)\left(2 + (x - 1)\left(1 + (x - 1)\left(-1 + (x - 1)(1 + 2(x - 2))\right)\right)\right).$
17. a) $2 + \frac{3}{10}x,$ b) $\frac{31}{10} + \frac{7}{10}x,$ c) $\frac{16}{5} + \frac{7}{10}x,$ d) $\frac{13}{5} + \frac{3}{5}x,$ e) $\frac{5}{4} - \frac{1}{2}x,$
 f) $\frac{11}{5} - \frac{2}{5}x,$ g) $1 + \frac{3}{10}x,$ h) $\frac{2}{5} + \frac{7}{10}x,$ i) $1 - x,$ j) $-\frac{7}{10}x.$
18. a) $x^2 + 2x - 1,$ b) $-2x^2 + x + 3,$ c) $-x^2 + 2x + 1,$ d) $x^2 - 3x - 5,$
 e) $-x^2 + 2x + 1,$ f) $x^2 - x - 1,$ g) $-x^2 + 2x - 1,$ h) $\frac{1}{2}x^2 - 2x - 1,$
 i) $x^2 + 2x + 1,$ j) $\frac{1}{2}x^2 - \frac{3}{2}x + 2.$

Testy ke kapitole 4

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Jsou dány hodnoty funkce $f(x)$ v tabulkových bodech x_i , $i = 0, \dots, n$. Funkci $f(x)$ chceme aproximovat funkcí $\varphi(x)$. Co musí splňovat tabulkové body x_i ?

Některé body mohou být stejné, ale musí v nich být různé funkční hodnoty.

Body musí být navzájem různé, tj. $x_i \neq x_j$ pro $i \neq j$, $i, j = 0, \dots, n$.

Body musí být v tabulce seřazeny vzestupně nebo sestupně.

Body x_i mohou být libovolné, ale musí být splněno $f(x_i) \neq f(x_j)$ pro všechna $i, j = 0, \dots, n$, $i \neq j$.

2. (1b.) Když pomocí aproximující funkce $\varphi(x)$ určené hodnotami v uzlových bodech x_0, x_1, \dots, x_n vypočteme funkční hodnotu v bodě x , který leží mimo interval $\text{Int}\langle x_0, x_n \rangle$, pak jsme provedli
- iteraci. regresi. extrapolaci. interpolaci.
3. (1b.) V případě, že nějakou funkci zadanou v $n + 1$ různých tabulkových bodech aproximujeme interpolačním polynomem $P(x)$, pak jeho stupeň
- je roven $n + 1$. je roven n .
- je nejvýše n . je nejméně n .
4. (1b.) Jsou dány dva různé tabulkové body, ve kterých je předepsaná stejná funkční hodnota. Pak grafem příslušného Lagrangeova interpolačního polynomu
- je parabola. je přímka rovnoběžná s osou x .
- je přímka se směrnicí různou od nuly. je přímka rovnoběžná s osou y .

7. (1b.) Funkce $f(x)$ je dána tabulkou. Jestliže chceme použít Newtonův interpolační polynom vpřed, pak
- uzly musí být uspořádané vzestupně a musí být ekvidistantní.
 - uzly musí být uspořádané vzestupně a vzdálenosti mezi sousedními uzly musí růst.
 - na uspořádání uzlů nezáleží, ale musí být ekvidistantní.
 - na uzly neklademe žádné požadavky.
8. (1b.) Při náhradě funkce $f(x)$ dané na intervalu pomocí interpolačního polynomu, který používá velký počet uzlových bodů,
- bývají chyby v blízkosti krajních bodů intervalu nulové.
 - mohou být chyby v blízkosti krajních bodů intervalu hodně velké.
 - se velikost chyb směrem ke krajním uzlovým bodům zmenšuje.
 - se velikost chyb s rostoucím počtem uzlů ustálí kolem jisté hodnoty.

9. (1b.) Doplňte správně konec věty: Poměrné diference

jsou jakousi náhradou derivací pro funkce, které jsou definované pouze na diskrétní množině.

závisí na pořadí uzlů, v nichž se počítají.

prvního řádu udávají koeficienty v Newtonově interpolačním polynomu.

nelze vypočítat v případě ekvidistantních uzlů.

10. (1b.) Hermitův interpolační polynom má čtyři uzly. Počet podmínek v nich zadaných je postupně 1, 2, 3 a 4. Vyberte správné tvrzení:

Funkční hodnota je předepsána v jednom uzlu.

Hodnoty první derivace jsou předepsány ve třech uzlech.

Hodnoty druhé derivace jsou předepsány ve třech uzlech.

Hodnoty třetí derivace jsou předepsány ve dvou uzlech.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (1b.) V $n + 1$ uzlových bodech x_i jsou předepsány hodnoty y_i . Co musí platit, aby dvojice $[x_i, y_i]$, $i = 0, \dots, n$, určovaly právě jeden interpolační polynom stupně nejvýše n ?

Uzlové body musí být navzájem různé, tj. $x_i \neq x_j$ pro $i \neq j$, $i, j = 0, \dots, n$.

Uzlové body musí být navzájem různé, tj. $x_i \neq x_j$ pro $i \neq j$, $i, j = 0, \dots, n$, a všechny hodnoty y_i musí být stejné.

V nejvýše dvou uzlových bodech mohou být předepsané stejné funkční hodnoty.

Pro všechna $i, j = 0, \dots, n$, $i \neq j$, musí být splněno $y_i \neq y_j$.

2. (1b.) Chceme aproximovat neznámou funkci polynomem, přičemž máme pouze tabulku diskrétních hodnot. Ve kterém případě je vhodnější použít metodu nejmenších čtverců místo interpolačního polynomu?

Když data nejsou rozsáhlá a víme o nich, že jsou přesná.

Když jsou uzly ekvidistantní.

Když jsou data získaná měřením.

Když máme málo dat.

3. (1b.) Vyberte předpis Newtonova interpolačního polynomu pro data z tabulky.

x_i	4,1	4,2
$f(x_i)$	1,95	2

$$N(x) = 1,95 + 2(x - 4,1).$$

$$N(x) = 4,1 + 4,2(x - 1,95).$$

$$N(x) = 1,95 + 0,5(x - 4,1).$$

$$N(x) = 1,95 + 5(x - 4,1).$$

4. (1b.) Které tvrzení o kubických interpolačních splajnech není pravdivé?

Málo se vlní a jejich grafy jsou křivky, které danými body procházejí hladce.

Jsou to po částech polynomiální funkce, které mají spojitou druhou derivaci.

K jejich nalezení se musí řešit nelineární soustavy rovnic, jejichž počet odpovídá počtu uzlů zmenšenému o jedničku.

Velmi dobře aproximují hladké funkce včetně jejich derivací.

5. (1b.) Doplňte: V případě aproximace funkce interpolačním polynomem na intervalu $\langle a, b \rangle$

kvalita aproximace obvykle roste s počtem uzlových bodů.

bývá aproximace nejlepší uprostřed intervalu $\langle a, b \rangle$.

bývá aproximace nejlepší v blízkosti krajních bodů intervalu $\langle a, b \rangle$.

6. (1b.) Doplňte: Pro hodnoty y_i předepsané v $n + 1$ různých tabulkových bodech x_i lze nalézt nekonečně mnoho různých interpolačních polynomů nejvýše n -tého stupně, jejichž grafy prochází body $[x_i, y_i]$.
- existuje právě jeden interpolační polynom nejvýše n -tého stupně, jehož graf prochází body $[x_i, y_i]$.
- existují alespoň dva různé interpolační polynomy nejvýše n -tého stupně, jejichž grafy prochází body $[x_i, y_i]$.
- existuje právě jeden interpolační polynom n -tého stupně, jehož graf prochází body $[x_i, y_i]$.
7. (1b.) Je dána tabulka hodnot $y_i = f(x_i)$, $i = 0, \dots, n$. Doplňte: V případě aproximace funkce $f(x)$ funkcí $\varphi(x) = c_0\varphi_0(x) + \dots + c_m\varphi_m(x)$ ve smyslu metody nejmenších čtverců můžeme koeficienty c_0, \dots, c_m zvolit libovolně.

musíme koeficienty c_0, \dots, c_m vybrat tak, aby výraz $\sum_{i=1}^n (y_i - \varphi(x_i))^2$ byl minimální.

musíme koeficienty c_0, \dots, c_m zvolit tak, aby výraz $\sum_{i=1}^n c_i^2$ byl minimální.

můžeme koeficienty c_0, \dots, c_m zvolit libovolně tak, aby odmocnina ze součtu jejich druhých mocnin byla rovna jedné.

8. (1b.) Je dáno $n + 1$ bodů $[x_i, y_i]$, přičemž pro jednu dvojici bodů platí $x_i = x_j$, ale $y_i \neq y_j$.

Pak těmito body lze proložit interpolační polynom nejvýše $(n - 1)$ -ního stupně.

Pak těmito body lze proložit interpolační polynom nejvýše $(n + 1)$ -ního stupně.

Pak těmito body lze proložit interpolační polynom nejvýše n -tého stupně.

Pak těmito body nelze proložit interpolační polynom.

9. (1b.) Je dána tabulka hodnot funkce $f(x)$ v $n + 1$ různých uzlových bodech uspořádaných podle velikosti. Pak kubickým interpolačním splajnem $S(x)$ nazýváme funkci, která je na každém intervalu $\langle x_i, x_{i+1} \rangle, i = 0, \dots, n,$

libovolným polynomem 3. stupně.

libovolným polynomem 3. stupně, přičemž $S(x_i) = f(x_i)$ pro $i = 0, \dots, n.$

libovolným polynomem 3. stupně a funkce $S(x)$ má ve všech bodech x_i první a druhé derivace.

libovolným polynomem 3. stupně, přičemž $S(x_i) = f(x_i)$ pro $i = 0, \dots, n$ a funkce $S(x)$ má ve všech bodech x_i první a druhé derivace.

10. (1b.) Hermitův interpolační polynom má čtyři uzly. Počet podmínek v nich zadaných je postupně 1, 2, 3 a 3. Vyberte správné tvrzení:

Funkční hodnoty jsou předepsány ve dvou uzlech.

Hodnoty první derivace jsou předepsány ve čtyřech uzlech.

Hodnoty druhé derivace jsou předepsány ve dvou uzlech.

Hodnoty třetí derivace jsou předepsány ve dvou uzlech.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (1b.) Neznámá funkce je zadaná tabulkou hodnot $[x_i, y_i]$. V čem spočívá rozdíl mezi interpolační aproximací této funkce a její aproximací metodou nejmenších čtverců?

Mezi metodami není žádný rozdíl, jde jen o různé postupy nalezení téže aproximující funkce.

Interpolační aproximace je přesnější, protože součet čtverců odchylek naměřených a aproximovaných hodnot je roven nule.

U interpolační aproximace graf aproximující funkce prochází body $[x_i, y_i]$, kdežto u aproximace metodou nejmenších čtverců těmito body procházet nemusí.

2. (1b.) Co je nevýhodou Lagrangeova interpolačního polynomu?

Pokud přidáme další tabulkový bod, přepočítaný polynom přestane v některých tabulkových bodech nabývat předepsanou hodnotu.

Pokud přidáme další tabulkový bod, musíme provést výpočty všech koeficientů znova.

Při jeho konstrukci někdy musíme dělit nulou.

Jeho nespojitost v bodech, které nejsou tabulkové.

3. (1b.) Je nutné, aby byly tabulkové body při konstrukci Newtonova interpolačního polynomu uspořádané podle velikosti?

Ne.

Ano.

4. (1b.) Pro speciální Lagrangeovy polynomy

$$L_i(x) = \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}$$

platí:

$$\text{st}L_i(x) = n + 1 \text{ a } L_i(x_i) = 1, L_i(x_j) = 0 \text{ pro } i \neq j.$$

$$\text{st}L_i(x) = n \text{ a } L_i(x_i) = 0, L_i(x_j) = 1 \text{ pro } i \neq j.$$

$$\text{st}L_i(x) = n \text{ a } L_i(x_i) = 1, L_i(x_j) = 0 \text{ pro } i \neq j.$$

$$\text{st}L_i(x) = n - 1 \text{ a } L_i(x_i) = 1, L_i(x_j) = 0 \text{ pro } i \neq j.$$

5. (1b.) Jsou dány tři různé tabulkové body, v nichž jsou předepsány různé funkční hodnoty. Proložíme-li jimi interpolační polynom, pak jeho grafem bude

přímka.

přímka nebo parabola.

přímka rovnoběžná s osou x .

parabola.

6. (1b.) Funkci $f(x)$ aproximujeme na intervalu $\langle x_0, x_n \rangle$ funkcí $\varphi(x)$. Co nazýváme chybou aproximace v bodě $x \in \langle x_0, x_n \rangle$?

Rozdíl mezi danou funkcí a její aproximující funkcí v posledním tabulkovém bodě x_n .

Rozdíl mezi danou funkcí a její aproximující funkcí v bodě x .

Absolutní hodnotu rozdílu mezi danou funkcí a její aproximující funkcí v bodě x .

7. (1b.) Funkce $f(x)$ je dána tabulkou s rostoucími ekvidistantními uzly. Newtonův interpolační polynom vzad využijeme pro výpočet aproximace funkčních hodnot funkce $f(x)$ v

netabulkových bodech, které jsou blízko středu tabulky.

netabulkových bodech, které jsou blízko začátku tabulky.

netabulkových bodech, které jsou blízko konci tabulky.

libovolných netabulkových bodech.

8. (1b.) Je dáno $n + 1$ rostoucích uzlových bodů x_0, x_1, \dots, x_n . Pak kubický splajn s defektem jedna je

libovolná funkce $S(x)$, která je na každém intervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, n - 1$, polynomem nejvýše 3. stupně a $S(x)$ má ve všech bodech x_i první derivace.

libovolná funkce $S(x)$, která je na každém intervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, n - 1$, polynomem nejvýše 3. stupně a $S(x)$ má ve všech bodech x_i druhé derivace.

libovolná funkce $S(x)$, která je na každém intervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, n - 1$, polynomem nejvýše 3. stupně a $S(x)$ má ve všech bodech x_i třetí derivace.

9. (1b.) Je dána tabulka hodnot $y_i = f(x_i)$, $i = 0, \dots, n$. Je možné, že při aproximaci funkce $f(x)$ polynomem $\varphi(x) = c_0 + c_1x + \dots + c_mx^m$ ve smyslu metody nejmenších

čtverců vyjde $\sum_{i=0}^n (y_i - \varphi(x_i))^2 = 0$?

Ano, např. vždy v případě, když $m = n$.

Ano, např. vždy v případě, když $m < n$.

Ano, tato možnost nastane, když funkce $f(x)$ bude mít ve dvou různých uzlových bodech stejné funkční hodnoty.

Ne, tato možnost nemůže nikdy nastat.

10. (1b.) Rozhodněte, které tvrzení je pravdivé.

Hermitův interpolační polynom je speciálním případem Newtonova interpolačního polynomu.

Hermitův interpolační polynom je speciálním případem Taylorova polynomu.

Hermitův interpolační polynom je zobecněním Lagrangeova interpolačního polynomu.

Hermitův interpolační polynom je speciálním případem Lagrangeova interpolačního polynomu.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 4

1. (1b.) Neznámá funkce je zadaná tabulkou hodnot. Chceme určit její hodnotu v nějakém bodě. Kdy hovoříme o interpolaci?

Když chceme určit hodnotu funkce v nějakém tabulkovém bodě.

V případě, kdy hledáme hodnotu funkce v bodě, který leží mezi dvěma tabulkovými body.

V případě, kdy hledáme hodnotu funkce v bodě, který leží vně intervalu určeného tabulkovými body.

V případě, že hledáme hodnotu funkce v nějakém tabulkovém bodě s danou přesností.

2. (1b.) Máme dva body $[x_0, f(x_0)]$ a $[x_1, f(x_1)]$, přičemž $x_0 \neq x_1$. Vyberte předpis jimi určeného Lagrangeova interpolačního polynomu.

$$L(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1).$$

$$L(x) = \frac{x - x_1}{x_0 - x_1} f(x_1) + \frac{x - x_0}{x_1 - x_0} f(x_0).$$

$$L(x) = \frac{x_1 - x}{x_0 - x_1} f(x_0) + \frac{x_0 - x}{x_1 - x_0} f(x_1).$$

$$L(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) - \frac{x - x_0}{x_1 - x_0} f(x_1).$$

3. (1b.) Platí obecně, že čím je vyšší stupeň interpolačního polynomu, tím menší se dopouštíme chyby?

Ne.

Ano.

4. (1b.) Je kubický interpolační splajn jednoznačně určen předepsáním funkčních hodnot v uzlových bodech?

Ano. Přidáním dalších podmínek bychom komplikovali konstrukci splajnu.

Ano. Přidáním dalších podmínek bude splajn přeurčen (tj. budeme mít více rovnic než neznámých parametrů).

Ne. Je potřeba přidat další dvě okrajové podmínky.

Ne. K jednoznačnému určení je potřeba přidat jednu okrajovou podmínku.

5. (1b.) Která metoda neslouží k nalezení Hermitova interpolačního polynomu?

Konstrukce Taylorova typu.

Konstrukce Lagrangeova typu.

Konstrukce Newtonova typu.

Metoda neurčitých koeficientů.

6. (1b.) V případě, že funkci $f(x)$ aproximujeme funkcí $\varphi(x) = ax^2 + bx + c$ ve smyslu nejmenších čtverců v uzlových bodech $x_i, i = 0, \dots, n$, pak funkci $\varphi(x)$ vybíráme tak, aby

její graf procházel všemi body $[x_i, f(x_i)], i = 0, \dots, n$.

se dala co nejsnadněji integrovat a derivovat.

se nám co nejvíc líbila.

byl minimalizován součet kvadrátů odchylek $f(x_i) - \varphi(x_i)$ přes všechny uzlové body.

7. (1b.) Poměrné diference 1. řádu jsou definovány vztahem (x_i jsou po dvou různé uzlové body):

$$f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i}. \quad f[x_i, x_{i+1}] = f[x_{i+1}] - f[x_i].$$

$$f[x_i] = \frac{f(x_i)}{x_i}. \quad \nabla f[x_i] = f[x_i] - f[x_{i-1}].$$

8. (1b.) Jsou-li předepsány hodnoty v $n + 1$ různých tabulkových bodech, pak Lagrangeův a Newtonův interpolační polynom pro tato data jsou dva různé polynomy.
- stupeň Lagrangeova i Newtonova interpolačního polynomu pro tato data je $n + 1$.
- Lagrangeův a Newtonův interpolační polynom pro tato data vyjadřují tentýž polynom, jenom jinak zapsaný.
- stupeň Lagrangeova a Newtonova interpolačního polynomu pro tato data nemusí být stejný.
9. (1b.) Jsou dány tři různé tabulkové body, v nichž jsou předepsány stejné funkční hodnoty. Proložíme-li jimi interpolační polynom, pak jeho grafem bude
- | | |
|--------------------------------|----------------------------------|
| parabola. | přímka se směrnicí rovnou jedné. |
| přímka rovnoběžná s osou x . | přímka rovnoběžná s osou y . |

10. (1b.) V případě, že funkci $f(x)$ nahradíme na intervalu $\langle a, b \rangle$ kubickým interpolačním splajnem a interpolačním polynomem s uzlovými body $a = x_0 < x_1 < \dots < x_n = b$, pak

splajn nemusí procházet všemi body $[x_i, f(x_i)]$, $i = 0, \dots, n$, a může tak lépe nahrazovat funkci $f(x)$.

splajn i polynom funkci dobře nahrazují uprostřed intervalu $\langle a, b \rangle$.

polynom je vždy lepší náhradou než splajn.

polynom je lepší náhradou v blízkosti koncových bodů intervalu.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Kapitola 5

Numerická derivace a integrace

Cíle

Po prostudování této kapitoly budete schopni vysvětlit:

- proč je nutné počítat určité integrály numericky a jaká je myšlenka jejich přibližného výpočtu,
- co jsou to kvadrurní formule,
- podle čeho posuzujeme přesnost kvadrurních formulí,
- co jsou to uzavřené a otevřené Newtonovy-Cotesovy kvadrurní formule a jak je dostaneme,

- jaký je rozdíl mezi jednoduchými a složenými kvadraturními formulemi,
- co jsou to Gaussovy kvadraturní formule a jaké mají přednosti,
- jaká je podstata Rombergovy kvadratury a kdy ji používáme,
- proč je nutné provádět numerickou derivaci a jaká je myšlenka jejího přibližného výpočtu,
- jaké problémy se objeví při numerickém derivování.

Otázky výpočtu derivace a jednoduchého určitého integrálu rovněž patří k základním úlohám řešeným v numerické matematice. V obou případech se používá stejný postup. Daná funkce se nahradí vhodnou aproximací a ta se derivuje resp. integruje. Mezi těmito úlohami je však zásadní rozdíl.

Určitý integrál totiž má jednu příjemnou vlastnost. Změníme-li málo hodnoty funkce $f(x)$, změní se málo i určitý integrál. Podrobněji, platí-li na intervalu $\langle a, b \rangle$ pro riemannovsky integrovatelné funkce $f(x)$ a $g(x)$, že $|f(x) - g(x)| < \varepsilon$, kde $\varepsilon > 0$, pak $|\int_a^b (f(x) - g(x)) dx| \leq \int_a^b |f(x) - g(x)| dx \leq \int_a^b \varepsilon dx = \varepsilon(b - a)$. Je-li tedy ε malé, je rozdíl mezi integrály $\int_a^b f(x) dx$ a $\int_a^b g(x) dx$ rovněž malý.

V případě derivace je však situace podstatně odlišná. Platí-li na intervalu $\langle a, b \rangle$ pro diferencovatelné funkce $f(x)$ a $g(x)$, že $|f(x) - g(x)| < \varepsilon$, kde $\varepsilon > 0$, pak i když je ε malé, může být rozdíl mezi derivacemi značně veliký. To není překvapivé, když si uvědomíme, že derivace $f'(x)$ je rovna směrnici ke grafu funkce $f(x)$ v bodě $[x, f(x)]$ a podobně pro funkci $g(x)$. Z blízkosti hodnot $f(x)$ a $g(x)$ rozhodně neplyne, že tečny k jejich grafům v bodech $[x, f(x)]$ a $[x, g(x)]$ mají přibližně stejný směr.

Např. pro funkce $f(x) = c$, $c \in \mathbb{R}$ (konstantní funkce) a $g(x) = c + \varepsilon \sin ax$, $\varepsilon > 0$, $a > 0$ platí $|f(x) - g(x)| = \varepsilon |\sin ax| \leq \varepsilon$. Pro jejich derivace však platí $|f'(x) - g'(x)| = |0 - \varepsilon a \cos ax| = |\varepsilon a \cos ax| \leq \varepsilon a$, přičemž poslední nerovnost přejde v rovnost, pokud je ax rovno celočíselnému násobku π . Proto i když je ε malé, pro velké číslo a bude absolutní hodnota rozdílu derivací v některých bodech veliká. Hodnoty funkce $g(x)$ se totiž od konstanty c liší jen málo (nejvýše o $\pm\varepsilon$), ale pro velké a velmi rychle oscilují (harmonická funkce $\varepsilon \sin ax$ má velkou frekvenci). Tudíž derivace $g'(x) = \varepsilon a \cos ax$ se od derivace $f'(x) = 0$ v některých bodech podstatně liší (až o $\pm\varepsilon a$). V animaci na obr. 5.1 je zvoleno $c = 1$, $\varepsilon = 0,1$ a a se mění od 1 do 30 s krokem 1. Pro $a \rightarrow \infty$ roste maximum εa absolutní hodnoty rozdílu mezi hodnotami derivací do nekonečna.

Obr. 5.1: Znázornění funkcí $f(x)$, $g(x)$, $f'(x)$ a $g'(x)$

U integrálů těchto funkcí je výsledek zcela jiný. Na libovolném intervalu $\langle p, q \rangle$ platí:

$$\int_p^q (f(x) - g(x)) dx = -\varepsilon \int_p^q \sin ax dx = \frac{\varepsilon}{a} [\cos ax]_p^q = \frac{\varepsilon(\cos aq - \cos ap)}{a}.$$

Tedy

$$\left| \int_p^q f(x) dx - \int_p^q g(x) dx \right| = \left| \frac{\varepsilon(\cos aq - \cos ap)}{a} \right| \leq \frac{2\varepsilon}{a} \rightarrow 0 \quad \text{pro } a \rightarrow \infty.$$

Tudíž s rostoucím a je naopak absolutní hodnota rozdílu mezi hodnotami integrálů čím dál menší a blíží se k nule pro $a \rightarrow \infty$.

Větší část této kapitoly je věnována numerické integraci, numerickou derivací se budeme zabývat pouze v posledních dvou oddílech.

5.1 Princip numerické integrace

Je-li $f(x)$ funkce riemannovsky integrovatelná na intervalu $\langle a, b \rangle$ a $F(x)$ je primitivní funkce k $f(x)$, tj. $F'(x) = f(x)$, platí dobře známá Newtonova-Leibnizova formule (viz [24, str. 100] nebo [25, str. 117])

$$\int_a^b f(x) dx = F(b) - F(a).$$

Její použití však nemusí být vždy dobře možné. Důvody jsou zejména následující:

- (1) Primitivní funkce $F(x)$ neleží ve třídě elementárních funkcí (nebo nějakých jiných vyšších transcendentních funkcí, u kterých umíme určit jejich funkční hodnoty), takže neumíme najít její vzorec, který by umožňoval vypočítat funkční hodnoty v bodech a a b .
- (2) Primitivní funkci $F(x)$ sice umíme najít, ale výpočet je velmi složitý.
- (3) Funkce $f(x)$ není zadaná vzorcem, známe pouze tabulku funkčních hodnot na konečné množině bodů (např. jsou to výsledky nějakého měření nebo algoritmu).

Jak již bylo řečeno v úvodu kapitoly, je-li $f(x) \approx g(x)$ na $\langle a, b \rangle$, platí $\int_a^b f(x) dx \approx \int_a^b g(x) dx$. Myšlenka přibližného výpočtu určitého integrálu je tedy zřejmá. Funkci $f(x)$ nahradíme na intervalu $\langle a, b \rangle$ nepříliš lišící se funkcí $g(x)$, kterou umíme integrovat, a integrál této náhražky prohlásíme za přibližnou hodnotu hledaného integrálu. Samozřejmě nás bude zajímat chyba, které se dopouštíme.

Náhradní funkcí může být např. interpolační polynom zadaný funkčními hodnotami funkce $f(x)$ ve vhodně zvolených bodech. Takový postup je použitelný i v případě, kdy nemáme analytické vyjádření funkce $f(x)$ a máme k dispozici jen konečnou tabulku jejích hodnot.

5.2 Kvadrurní formule

Jedna z možných definic Riemannova¹ integrálu je založena na jistém typu limity integrálních součtů. Připomeňme jejich definici.

Zvolíme dělení $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$ základního intervalu a v každém podintervalu zvolíme reprezentant $\xi_i \in \langle x_{i-1}, x_i \rangle$. Integrální součet pak definujeme vztahem

$$f(\xi_1)(x_1 - x_0) + f(\xi_2)(x_2 - x_1) + \dots + f(\xi_n)(x_n - x_{n-1}).$$

Jde vlastně o součet funkčních hodnot v nějakých bodech, které jsou násobené jistými čísly (v tomto případě délkami podintervalů). Při zvětšování počtu podintervalů a neomezeném zkracování jejich délek se bude integrální součet neomezeně přibližovat k hodnotě integrálu $\int_a^b f(x) dx$.

¹**Georg Friedrich Bernhard Riemann** (1826–1866) (čti ríman) — vynikající německý matematik. Zabýval se teorií funkcí, geometrií, matematickou a teoretickou fyzikou a diferenciálními rovnicemi. Je považován za jednoho z největších matematiků všech dob. Jeho tzv. *Riemannova hypotéza* o rozložení nulových bodů ζ -funkce (čti dzéta-funkce) je dodnes nevyřešena a patří k nejslavnějším matematickým problémům.

Pokusíme se najít přibližnou hodnotu integrálu

$$I(f) = \int_a^b f(x) dx$$

v podobném tvaru. Zvolíme $n + 1$ bodů x_i , $a \leq x_0 < x_1 < \dots < x_{n-1} < x_n \leq b$, kde $n \geq 0$. Koncové body a a b mezi nimi mohou být ale nemusí. Dále zvolíme čísla A_i , $i = 0, 1, \dots, n$. Výraz

$$Q(f) = A_0 f(x_0) + A_1 f(x_1) + \dots + A_n f(x_n) \quad (5.1)$$

nazýváme *kvadrurní formule*. Body x_i , $i = 0, 1, \dots, n$, nazýváme *uzly* a čísla A_i , $i = 0, 1, \dots, n$, nazýváme *koeficienty* kvadrurní formule (5.1). Obecně bude platit pouze $I(f) \approx Q(f)$. Označme jejich rozdíl $R(f)$, tedy

$$I(f) = Q(f) + R(f). \quad (5.2)$$

Hodnota $R(f)$ představuje *chybu kvadrurní formule*. Počet uzlů n nebude v praxi příliš velký. Otázkou pak je, jak při daném n zvolit uzly a koeficienty, aby chyba $R(f)$ byla co nejmenší.

Kvalitu kvadraturní formule můžeme rovněž posuzovat podle toho, zda je přesná pro nějaké jednoduché funkce. Jako testovací funkce zvolíme mocniny x^k , $k = 0, 1, 2, \dots$. Jejich grafy jsou postupně přímka rovnoběžná s osou x , přímka, parabola, atd.

Definice 5.1 Řekneme, že kvadraturní formule

$$Q(f) = \sum_{i=0}^n A_i f(x_i)$$

má stupeň přesnosti N , jestliže platí

$$R(1) = R(x) = \dots = R(x^N) = 0, \quad R(x^{N+1}) \neq 0.$$

Kvadraturní formule (5.1) je zřejmě lineární: jsou-li $f(x)$ a $g(x)$ funkce a α, β konstanty, platí $Q(\alpha f + \beta g) = \alpha Q(f) + \beta Q(g)$. Protože integrál je rovněž lineární (integrál ze součtu je součet integrálů a multiplikační konstantu můžeme vytknout), kvadraturní formule mající stupeň přesnosti N je zcela přesná (při počítání bez zaokrouhlování) pro libovolný polynom stupně nejvýše N . Pro polynomy stupně $N + 1$ však už

přesná není.

Následující věta dává částečně odpověď na otázku, jak velký stupeň přesnosti lze dosáhnout při daném počtu uzlů.

Věta 5.2 *Kvadrurní formule používající $n + 1$ bodů může mít stupeň přesnosti nejvýše $2n + 1$.*

Důkaz. Pripusťme, že kvadrurní formule (5.1) má stupeň přesnosti $2n + 2$. Označme $\omega(x) = (x - x_0)^2 \cdots (x - x_n)^2$. Pak $\omega(x)$ je polynom stupně $2n + 2$, takže by mělo pro chybu kvadrurní formule platit $R(\omega) = 0$. Přitom $\omega(x_i) = 0$, $i = 0, \dots, n$, takže $Q(\omega) = 0$. Tudíž

$$0 = R(\omega) = I(\omega) - Q(\omega) = \int_a^b \omega(x) dx,$$

což je spor, protože funkce $\omega(x)$ je na intervalu $\langle a, b \rangle$ kladná s výjimkou $n + 1$ uzlových bodů, takže nemůže mít nulový integrál. \square

Otázkou je, zda kvadrurní formule užívající $n + 1$ uzlů může mít stupeň přesnosti $2n + 1$. Později uvidíme, že to možné je.

Podívejme se nyní, co bude platit pro kvadraturní formuli, kterou získáme integrací interpolačního polynomu. O funkci $f(x)$ předpokládáme, že je dostatečně hladká (má derivace aspoň do řádu $n + 1$).

Věta 5.3 *Kvadraturní formule získaná integrací interpolačního polynomu $L(x)$ určeného body $[x_i, f(x_i)]$, $i = 0, 1, \dots, n$, má stupeň přesnosti nejméně n .*

Důkaz. Uvažujme Lagrangeův interpolační polynom ve tvaru (4.3)

$$L(x) = \sum_{i=0}^n f(x_i) L_i(x).$$

Označme $\omega(x) = (x - x_0) \cdots (x - x_n)$. Podle věty 4.4 platí

$$f(x) = L(x) + \omega(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}, \quad \text{kde } \xi = \xi(x).$$

Integrací předchozího vztahu dostaneme

$$\int_a^b f(x) dx = \sum_{i=0}^n f(x_i) \int_a^b L_i(x) dx + \frac{1}{(n+1)!} \int_a^b \omega(x) f^{(n+1)}(\xi) dx.$$

Označme

$$A_i = \int_a^b L_i(x) dx, \quad i = 0, 1, \dots, n.$$

Pak

$$I(f) = Q(f) + R(f),$$

kde

$$Q(f) = \sum_{i=0}^n A_i f(x_i) \quad \text{a} \quad R(f) = \frac{1}{(n+1)!} \int_a^b \omega(x) f^{(n+1)}(\xi) dx.$$

Protože $(n+1)$ -ní derivace polynomu stupně nejvýše n je identicky rovna nule, je pro $f(x) = x^k$, $k = 0, 1, \dots, n$, integrand ve vzorci pro chybu $R(f)$ roven nule pro libovolné $x \in \langle a, b \rangle$, takže platí $R(1) = R(x) = \dots = R(x^n) = 0$, tudíž stupeň přesnosti kvadraturní formule $Q(f)$ je nejméně n . \square

Snadno se ověří, že koeficienty A_i v kvadraturní formuli (5.1), která má stupeň přesnosti aspoň n , jsou určeny jednoznačně. Dosadíme-li totiž do vztahu (5.2) za $f(x)$ postupně mocniny x^k , $k = 0, 1, \dots, n$, obdržíme pro tyto koeficienty čtvercovou soustavu

Poznámka 5.4

- a) Snadno lze ukázat, že když jsou uzly kvadraturní formule (5.1) získané integrací interpolačního polynomu rozloženy souměrně vzhledem ke středu intervalu $\langle a, b \rangle$, tj. bodu $\frac{a+b}{2}$, jsou koeficienty odpovídající souměrným uzlům stejné, tedy $A_i = A_{n-i}$, $i = 0, 1, \dots, n$ (viz [3, str. 97]). Takové formule nazveme *symetrické*.
- b) Symetrické kvadraturní formule z bodu a), které používají lichý počet uzlů $n + 1$ (tedy n je sudé a střed $\frac{a+b}{2}$ je jedním z uzlů), mají stupeň přesnosti aspoň $n + 1$, tudíž o jedničku víc než zaručuje v obecném případě věta 5.3 (viz [3, str. 98], [22, str. 230]).

V následujících oddílech se budeme zabývat nalezením konkrétních kvadraturních formulí. Všímneme si dvou typů:

- a) Vyjdeme z ekvidistantního dělení intervalu $\langle a, b \rangle$ na n podintervalů. Uzly tedy budou dány.
- b) Uzly a koeficienty zvolíme tak, aby při daném počtu uzlů byl dosažen maximální možný stupeň přesnosti.

Poznámka 5.5 Někdy se studuje výpočet obecnějších typů integrálů tvaru $\int_a^b w(x) f(x) dx$, kde $w(x) > 0$ je tzv. *váhová funkce*. To je vhodné, pokud je třeba určit hodnoty více integrálů, jejichž

integrandy mají společnou část $w(x)$ a mění se jen $f(x)$. Kvadrurní formule (5.1) se nemění, váhová funkce se „zahrne“ do koeficientů. Rovněž se uvažují případy nevlastních integrálů prvního druhu, kdy $a = -\infty$ nebo $b = \infty$, nebo druhého druhu, kdy integrand $w(x)f(x)$ je neohraničená funkce. Viz např. [22, 52]. My se omezíme jen na vlastní integrály a případ $w(x) = 1$.

5.3 Newtonovy-Cotesovy kvadrurní formule

Interval $\langle a, b \rangle$ rozdělíme na n podintervalů stejné délky $h > 0$. Tedy $h = \frac{b-a}{n}$. Položíme $x_0 = a$, $x_1 = x_0 + h$, $x_2 = x_1 + h$, ..., $x_n = x_0 + nh = b$. Dále pro stručnost označíme $y_i = f(x_i)$, $i = 0, 1, \dots, n$.

Rozlišíme dva případy:

- 1) Krajní body $x_0 = a$ a $x_n = b$ zahrneme mezi uzly. Počet uzlů tedy bude $n + 1$. Výsledné formule se nazývají *uzavřené*.
- 2) Krajní body $x_0 = a$ a $x_n = b$ nezahrneme mezi uzly. Počet uzlů tedy bude $n - 1$. Výsledné formule se nazývají *otevřené*.

Kvadrurní formule získáme integrací interpolačního polynomu určeného hodno-

tami funkce $f(x)$ ve zvolených uzlových bodech. Nazývají se *Newtonovy*¹-*Cotesovy*² kvadrurní formule.

5.3.1 Uzavřené Newtonovy-Cotesovy kvadrurní formule

Za uzly zvolíme $n + 1$ bodů $x_0 < x_1 < \dots < x_n$, $n \geq 1$. Podle (4.3) má Lagrangeův interpolační polynom tvar

$$L(x) = y_0 L_0(x) + \dots + y_n L_n(x), \quad \text{kde} \quad L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}, \quad i = 0, \dots, n.$$

Integrací na intervalu $\langle a, b \rangle$ dostaneme (viz důkaz věty 5.3)

$$\int_a^b f(x) dx = \sum_{i=0}^n y_i \int_a^b L_i(x) dx + R(f) = \sum_{i=0}^n A_i y_i + R(f),$$

¹Viz str.318

²**Roger Cotes** (1682–1716) (čti kouts) — anglický astronom a matematik.

kde

$$A_i = \int_a^b L_i(x) dx, \quad i = 0, 1, \dots, n.$$

Nyní zavedeme novou proměnnou t vztahem $x = a + th$. Po dosazení do $L_i(x)$ vyjde

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{t - j}{i - j} = \varphi_i(t).$$

Odtud po substituci dostaneme

$$A_i = \int_a^b L_i(x) dx = \left| \begin{array}{l} x = a + th \\ dx = h dt \\ a \rightsquigarrow 0, b \rightsquigarrow n \end{array} \right| = h \int_0^n \varphi_i(t) dt. \quad (5.4)$$

Z předchozího vztahu je vidět, že koeficienty A_i nezávisí na funkci $f(x)$ ani na mezích a a b . Navíc, vzhledem k tomu, že $\varphi_i(t)$ jsou polynomy s racionálními koeficienty a meze posledního integrálu jsou celočíselné, jsou A_i racionální čísla. Dále podle poznámky 5.4 a) platí $A_i = A_{n-i}$, $i = 0, 1, \dots, n$.

Nyní si všimneme chyby kvadrurních formulí. O funkci $f(x)$ předpokládáme, že je dostatečně hladká (má spojitě derivace dostatečně vysokého řádu).

V důkazu věty 5.3 jsme dostali, že

$$R(f) = \frac{1}{(n+1)!} \int_a^b \omega(x) f^{(n+1)}(\xi) dx,$$

kde $\omega(x) = (x - x_0) \cdots (x - x_n)$. Lze ukázat (viz [51, str. 157, 165], [52, str. 147]), že

$$R(f) = h^{p+1} K f^{(p)}(\xi), \quad \text{kde } a < \xi < b. \quad (5.5)$$

Přitom konstanty K a p závisí jen na n , ne na funkci $f(x)$ ($p = n + 1$ pro $n + 1$ sudé, $p = n + 2$ pro $n + 1$ liché; $n + 1$ je počet uzlů).

Nyní uvedeme Newtonovy-Cotesovy uzavřené formule pro $n = 1, 2, 3, 4$. Nejprve určíme pomocí vztahu (5.4) koeficienty A_i . Využijeme přitom symetrie. Označme $\alpha_i = A_i/h$. Postupně dostaneme:

$$n = 1 \quad \alpha_0 = \alpha_1 = \int_0^1 \varphi_0(t) = \int_0^1 \frac{t-1}{-1} dt = \frac{1}{2},$$

$$n = 2 \quad \alpha_0 = \alpha_2 = \int_0^2 \varphi_0(t) = \int_0^2 \frac{(t-1)(t-2)}{(-1) \cdot (-2)} dt = \frac{1}{3},$$

$$\alpha_1 = \int_0^2 \varphi_1(t) = \int_0^2 \frac{t(t-2)}{1 \cdot (-1)} dt = \frac{4}{3},$$

$$n = 3 \quad \alpha_0 = \alpha_3 = \int_0^3 \varphi_0(t) = \int_0^3 \frac{(t-1)(t-2)(t-3)}{(-1) \cdot (-2) \cdot (-3)} dt = \frac{3}{8},$$

$$\alpha_1 = \alpha_2 = \int_0^3 \varphi_1(t) = \int_0^3 \frac{t(t-2)(t-3)}{1 \cdot (-1) \cdot (-2)} dt = \frac{9}{8},$$

$$n = 4 \quad \alpha_0 = \alpha_4 = \int_0^4 \varphi_0(t) = \int_0^4 \frac{(t-1)(t-2)(t-3)(t-4)}{(-1) \cdot (-2) \cdot (-3) \cdot (-4)} dt = \frac{14}{45},$$

$$\alpha_1 = \alpha_3 = \int_0^4 \varphi_1(t) = \int_0^4 \frac{t(t-2)(t-3)(t-4)}{1 \cdot (-1) \cdot (-2) \cdot (-3)} dt = \frac{64}{45},$$

$$\alpha_2 = \int_0^4 \varphi_2(t) = \int_0^4 \frac{t(t-1)(t-3)(t-4)}{2 \cdot 1 \cdot (-1) \cdot (-2)} dt = \frac{8}{15}.$$

Nyní uvedeme příslušné kvadraturní formule. Odvození vzorců pro chyby lze nalézt např. v [22, 52]. Platí $A_i = \alpha_i h$, kde $b - a = nh$. Po drobných úpravách vyjde (ve všech vzorcích platí $a < \xi < b$, počet uzlů je $n + 1$):

$$I(f) = \frac{b-a}{2}(y_0 + y_1) - \frac{1}{12} h^3 f''(\xi), \quad (5.6)$$

(lichoběžníkové pravidlo, $n = 1$)

$$I(f) = \frac{b-a}{6}(y_0 + 4y_1 + y_2) - \frac{1}{90} h^5 f^{(4)}(\xi), \quad (5.7)$$

(Simpsonovo¹ pravidlo, $n = 2$)

$$I(f) = \frac{b-a}{8}(y_0 + 3y_1 + 3y_2 + y_3) - \frac{3}{80} h^5 f^{(4)}(\xi), \quad (5.8)$$

(Simpsonovo 3/8 pravidlo, $n = 3$)

$$I(f) = \frac{b-a}{90}(7y_0 + 32y_1 + 12y_2 + 32y_3 + 7y_4) - \frac{8}{945} h^7 f^{(6)}(\xi). \quad (5.9)$$

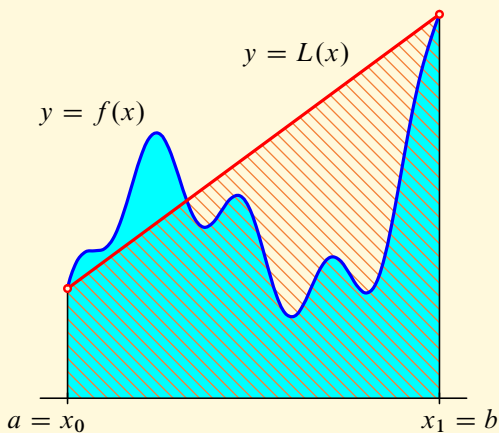
(Boolovo² pravidlo, $n = 4$)

¹**Thomas Simpson** (1710–1761) — anglický matematik.

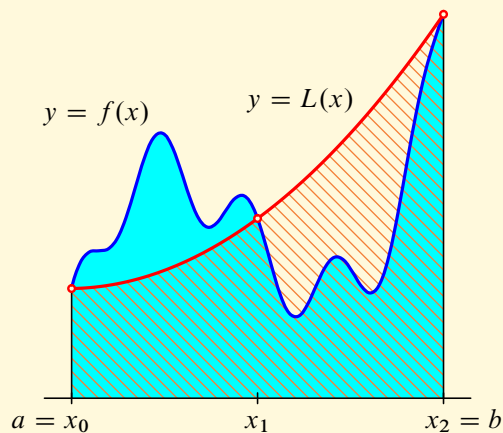
²**George Boole** (1815–1864) (čti bul) — anglický matematik. Významně přispěl k rozvoji matematické logiky.

Vzorce pro chyby v předchozích čtyřech kvadrurních formulích obsahují hodnotu derivace funkce $f(x)$ řádu $n + 1$ resp. $n + 2$ v neznámém bodě ξ . Je-li $f(x)$ polynomem stupně nejvýše n resp. $n + 1$, je $(n + 1)$ -ní resp. $(n + 2)$ -há derivace funkce $f(x)$ identicky nulová, na hodnotě ξ nezáleží a chyba je rovna nule. Je-li tudíž počet uzlů $n + 1$ sudé číslo (vzorce (5.6) a (5.8)), je stupeň přesnosti roven n , zatímco když je počet uzlů $n + 1$ liché číslo (vzorce (5.7) a (5.9)), je stupeň přesnosti roven $n + 1$. To je ve shodě s větou 5.3 a poznámkou 5.4 b). Stejně vlastnosti, pokud jde o stupeň přesnosti, mají uzavřené Newtonovy-Cotesovy formule i pro $n > 4$.

Grafický význam vzorců je znázorněn na obr. 5.2 a 5.3. Pro $n = 1$ jde o náhradu přímkou. Nabývá-li funkce $f(x)$ pouze kladné hodnoty, udává tato kvadrurní formule obsah lichoběžníku. Odtud pochází její název *lichoběžníkové pravidlo*. U *Simpsonova pravidla* jde o náhradu parabolou, u zbývajících dvou, *Simpsonova 3/8 pravidla* resp. *Boolova pravidla*, pak jde o náhradu polynomem stupně tři resp. čtyři.

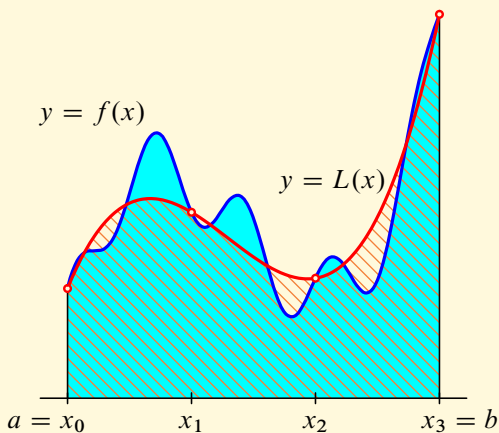


a) Lichoběžníkové pravidlo

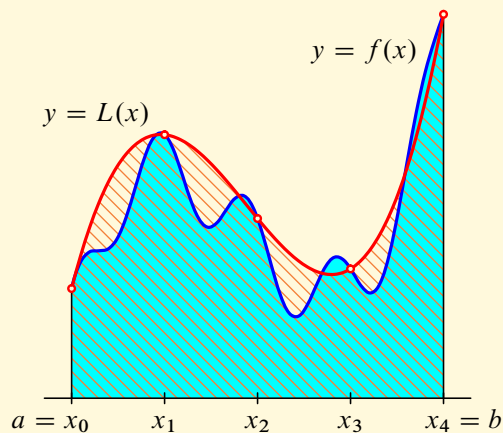


b) Simpsonovo pravidlo

Obr. 5.2: Newtonovy-Cotesovy uzavřené formule — část 1



a) Simpsonovo 3/8 pravidlo



b) Boolovo pravidlo

Obr. 5.3: Newtonovy-Cotesovy uzavřené formule — část 2

5.3.2 Otevřené Newtonovy-Cotesovy kvadrurní formule

Budeme postupovat obdobně jako u uzavřených Newtonových-Cotesových formulí. Za uzly zvolíme $n - 1$ bodů $x_1 < x_2 < \dots < x_{n-1}$, $n \geq 2$. Podle (4.3) má Lagrangeův interpolační polynom tvar

$$L(x) = y_1 L_1(x) + \dots + y_{n-1} L_{n-1}(x),$$

kde

$$L_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \frac{x - x_j}{x_i - x_j}, \quad i = 1, \dots, n-1, \quad (\text{pro } n = 2 \text{ je } L_1(x) = 1).$$

Integrací na intervalu $\langle a, b \rangle$ dostaneme (viz důkaz věty 5.3)

$$\int_a^b f(x) dx = \sum_{i=1}^{n-1} y_i \int_a^b L_i(x) dx + R(f) = \sum_{i=1}^{n-1} A_i y_i + R(f),$$

kde

$$A_i = \int_a^b L_i(x) dx, \quad i = 1, 2, \dots, n-1.$$

Nyní zavedeme novou proměnnou t vztahem $x = a + th$. Po dosazení do $L_i(x)$ vyjde

$$L_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \frac{x - x_j}{x_i - x_j} = \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \frac{t - j}{i - j} = \psi_i(t), \quad (\text{pro } n = 2 \text{ je } \psi_1(x) = 1).$$

Odtud po substituci dostaneme

$$A_i = \int_a^b L_i(x) dx = \left| \begin{array}{l} x = a + th \\ dx = h dt \\ a \rightsquigarrow 0, b \rightsquigarrow n \end{array} \right| = h \int_0^n \psi_i(t) dt. \quad (5.10)$$

Z předchozího vztahu je vidět, že koeficienty A_i nezávisí na funkci $f(x)$ ani na mezích a a b . Navíc, vzhledem k tomu, že $\psi_i(t)$ jsou polynomy s racionálními koeficienty a meze posledního integrálu jsou celočíselné, jsou A_i racionální čísla. Dále podle poznámky 5.4 a) platí $A_i = A_{n-i}$, $i = 1, 2, \dots, n-1$.

Nyní si všimneme chyby kvadraturních formulí. V důkazu věty 5.3 jsme dostali (zaměníme $n + 1$ za $n - 1$ a označíme $\omega(x) = (x - x_1) \cdots (x - x_{n-1})$), že

$$R(f) = \frac{1}{(n-1)!} \int_a^b \omega(x) f^{(n-1)}(\xi) dx.$$

Platí i (5.5) ($p = n - 1$ pro $n - 1$ sudé, $p = n$ pro $n - 1$ liché; $n - 1$ je počet uzlů).

Nyní uvedeme Newtonovy-Cotesovy otevřené formule pro $n = 2, 3, 4, 5$. O funkci $f(x)$ předpokládáme, že je dostatečně hladká (má spojité derivace dostatečně vysokého řádu).

Nejprve určíme pomocí vztahu (5.10) koeficienty A_i . Využijeme přitom symetrie. Označme $\alpha_i = A_i/h$. Postupně dostaneme:

$$n = 2 \quad \alpha_1 = \int_0^2 \psi_1(t) dt = \int_0^2 1 dt = 2,$$

$$n = 3 \quad \alpha_1 = \alpha_2 = \int_0^3 \psi_1(t) dt = \int_0^3 \frac{t-2}{-1} dt = \frac{3}{2},$$

$$n = 4 \quad \alpha_1 = \alpha_3 = \int_0^4 \psi_1(t) = \int_0^4 \frac{(t-2)(t-3)}{(-1) \cdot (-2)} dt = \frac{8}{3},$$

$$\alpha_2 = \int_0^4 \psi_2(t) = \int_0^4 \frac{(t-1)(t-3)}{1 \cdot (-1)} dt = -\frac{4}{3},$$

$$n = 5 \quad \alpha_1 = \alpha_4 = \int_0^5 \psi_1(t) = \int_0^5 \frac{(t-2)(t-3)(t-4)}{(-1) \cdot (-2) \cdot (-3)} dt = \frac{55}{24},$$

$$\alpha_2 = \alpha_3 = \int_0^5 \psi_2(t) = \int_0^5 \frac{(t-1)(t-3)(t-4)}{1 \cdot (-1) \cdot (-2)} dt = \frac{5}{24}.$$

Nyní uvedeme příslušné kvadraturní formule. Odvození vzorců pro chyby lze nalézt např. v [22, 52]. Platí $A_i = \alpha_i h$, kde $b - a = nh$. Po drobných úpravách vyjde (ve všech vzorcích platí $a < \xi < b$, počet uzlů je $n - 1$):

$$I(f) = (b - a)y_1 + \frac{1}{3} h^3 f''(\xi), \quad (5.11)$$

(obdélníkové pravidlo, $n = 2$)

$$I(f) = \frac{b-a}{2}(y_1 + y_2) + \frac{1}{4} h^3 f''(\xi), \quad (5.12)$$

(otevřené lichoběžníkové pravidlo, $n = 3$)

$$I(f) = \frac{b-a}{3}(2y_1 - y_2 + 2y_3) + \frac{14}{45} h^5 f^{(4)}(\xi), \quad (5.13)$$

(Milnovo¹ pravidlo, $n = 4$)

$$I(f) = \frac{b-a}{24}(11y_1 + y_2 + y_3 + 11y_4) + \frac{95}{144} h^5 f^{(4)}(\xi). \quad (5.14)$$

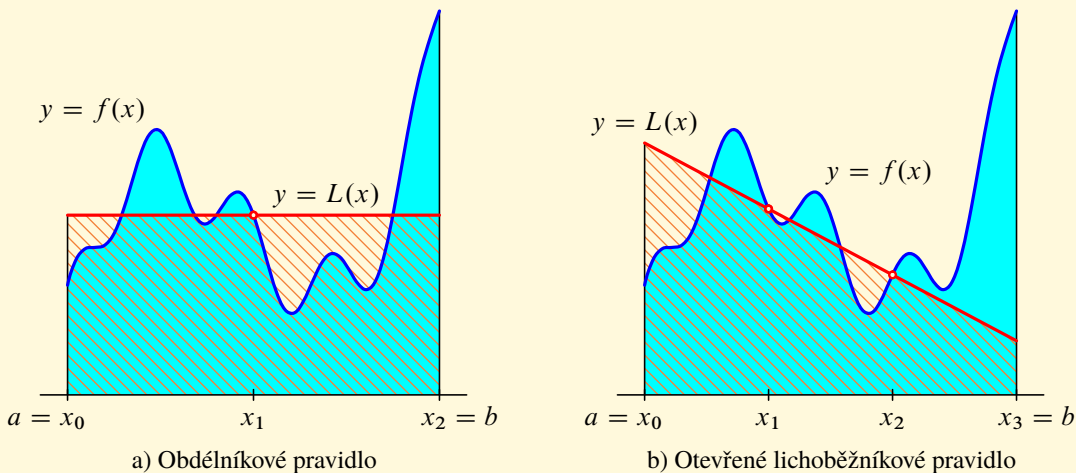
($n = 5$)

Vzorce pro chyby v předchozích čtyřech kvadraturních formulích obsahují hodnotu derivace funkce $f(x)$ řádu $n - 1$ resp. n v neznámém bodě ξ . Je-li $f(x)$ polynomem

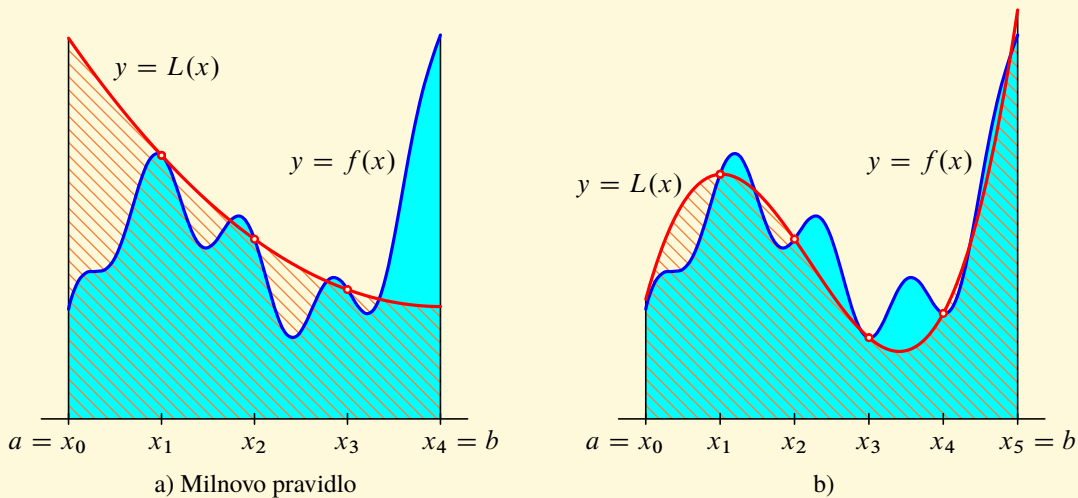
¹**William Edmund Milne** (1890–1971) (čti miln) — americký matematik.

stupně nejvýše $n - 2$ resp. $n - 1$, je $(n - 1)$ -ní resp. n -tá derivace funkce $f(x)$ identicky nulová, na hodnotě ξ nezáleží a chyba je rovna nule. Je-li tudíž počet uzlů $n - 1$ sudé číslo (vzorce (5.12) a (5.14)), je stupeň přesnosti roven $n - 2$, zatímco když je počet uzlů $n - 1$ liché číslo (vzorce (5.11) a (5.13)), je stupeň přesnosti roven $n - 1$. To je opět ve shodě s větou 5.3 a poznámkou 5.4 b). Stejně vlastnosti, pokud jde o stupeň přesnosti, mají otevřené Newtonovy-Cotesovy formule i pro $n > 5$.

Grafický význam vzorců je znázorněn na obr. 5.4 a 5.5. Pro $n = 2$ jde o náhradu přímkou rovnoběžnou s osou x . Nabývá-li funkce $f(x)$ pouze kladné hodnoty, udává tato kvadrurní formule obsah obdélníku. Odtud pochází její název *obdélníkové pravidlo*. Pro $n = 3$ jde o náhradu přímkou. Pokud nabývá Lagrangeův polynom $L(x)$, kterým nahrazujeme funkci $f(x)$, pouze kladné hodnoty, udává tato kvadrurní formule obsah lichoběžníku. Odtud pochází její název *otevřené lichoběžníkové pravidlo*. U *Milnova pravidla* jde o náhradu parabolou, u posledního pravidla, které nemá speciální název, pak jde o náhradu polynomem stupně tři.



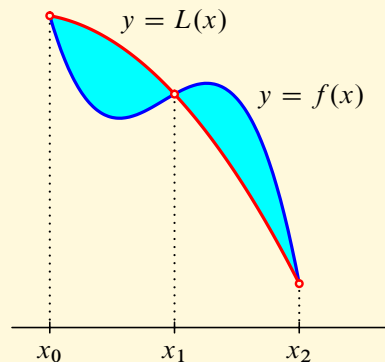
Obr. 5.4: Newtonovy-Cotesovy otevřené formule — část 1



Obr. 5.5: Newtonovy-Cotesovy otevřené formule — část 2

Nezaměňujte otevřené lichoběžníkové pravidlo (5.11) s lichoběžníkovým pravidlem (5.6). Porovnejte rozdíl na obrázcích 5.2 a) a 5.4 b). Rovněž h má ve vzorcích (5.6) a (5.12) jiný význam. V prvním je $h = b - a$, ve druhém pak $h = (b - a)/3$. Podobné rozdíly jsou mezi Simpsonovým pravidlem (5.7) a Milnovým pravidlem (5.13), u kterých je funkce $f(x)$ nahrazována parabolou.

V předchozím výkladu jsme uvedli, že pro lichý počet $n + 1$ použitých uzlů je stupeň přesnosti Newtonovy-Cotesovy formule roven $n + 1$, tj. o jedničku víc, než bychom očekávali. Vysvětlíme si tuto skutečnost na Simpsonově pravidle (5.7). Funkce $f(x)$ na obr. 5.6 představuje obecný kubický polynom. Ten nahradíme interpolačním polynomem $L(x)$, který nabývá v bodech x_0 , x_1 a x_2 stejné hodnoty jako $f(x)$. Přitom $L(x)$ je kvadratický, takže jeho grafem je parabola, která rozhodně nesplývá s grafem $f(x)$. Protože stupeň přesnosti Simpsonova pravidla je tři, je chyba nulová a platí



Obr. 5.6

$$\int_{x_0}^{x_2} [f(x) - L(x)] dx = 0.$$

Odtud dostaneme po rozdělení intervalu $\langle x_0, x_2 \rangle$ na $\langle x_0, x_1 \rangle$ a $\langle x_1, x_2 \rangle$ a jednoduché úpravě, že

$$\int_{x_1}^{x_2} [f(x) - L(x)] dx = \int_{x_0}^{x_1} [L(x) - f(x)] dx.$$

To znamená, že plochy mezi grafy funkcí $f(x)$ a $L(x)$ na intervalech $\langle x_0, x_1 \rangle$ a $\langle x_1, x_2 \rangle$ jsou stejné. Tato skutečnost je příčinou zvýšené přesnosti. Rozhodně není správná představa, kterou studenti někdy mají, že grafy $f(x)$ a $L(x)$, tedy parabola a kubická křivka, na $\langle x_0, x_2 \rangle$ splývají. Obdobně je tomu i u jiných Newtonových-Cotesových formulí využívajících lichý počet uzlů. Rozmyslete si, jak vypadá analogie obr. 5.6 pro obdélníkové pravidlo.

5.3.3 Složené Newtonovy-Cotesovy kvadrurní formule

Uvažujme kvadrurní formule, které jsou získané integrací interpolačního polynomu. Předpokládejme, že interval $\langle a, b \rangle$ je dlouhý. Zvolíme-li malý počet uzlů, interpolační polynom pravděpodobně nebude dobře aproximovat integrand $f(x)$ a chyba může být velká. Zvolíme-li velký počet uzlů, může interpolační polynom mezi uzly oscilovat (srovnejte příklad 4.5), rovněž nebude dobře aproximovat $f(x)$ a chyba bude opět velká.

Proto obvykle postupujeme v podobných případech jinak. Výchozí interval $\langle a, b \rangle$ rozdělíme na menší podintervaly a na každém z nich použijeme základní kvadrurní formuli s nepřilíš velkým počtem uzlových bodů. Výsledky sečteme. Protože podintervaly jsou krátké, dá se očekávat, že chyby na každém z nich nebudou velké a rovněž celková chyba, která je jejich součtem, nebude přilíš velká.

My se omezíme na případ, kdy podintervaly budou stejně dlouhé. Na každém z nich použijeme analogickou základní Newtonovu-Cotesovu kvadrurní formuli a dostaneme tak odpovídající *složenou kvadrurní formuli*.

Poznámka 5.6 Někdy funkce $f(x)$ na částech intervalu $\langle a, b \rangle$ výrazně osciluje a na jiných se téměř nemění. Pak je vhodné volit v místech rychlých změn kratší podintervaly a v místech, kde je integrand téměř konstantní, delší intervaly. Na těchto úvahách jsou založeny tzv. *adaptivní kvadraturní formule*. Podrobněji viz např. [22, str. 264].

K vytvoření složené formule lze v podstatě použít libovolnou základní formuli. Obvykle se však pro tento účel nevolí formule používající velký počet uzlů. Rovněž otevřené Newtonovy-Cotesovy formule nejsou příliš vhodné, protože nepoužívají krajní body intervalu, což je nevýhodné, když chceme např. kvůli zpřesnění výsledku zdvojnásobit počet základních intervalů. Uvedeme tři nejčastější složené vzorce, které vycházejí z obdélníkového, lichoběžníkového nebo Simpsonova pravidla. O funkci $f(x)$ budeme předpokládat, že je na intervalu $\langle a, b \rangle$ dostatečně hladká. Tento předpoklad je nutný k odvození tvaru zbytku.

Složené obdélníkové pravidlo

Protože obdélníkové pravidlo (5.11) dělí základní interval na dvě části, rozdělíme celý interval $\langle a, b \rangle$ na sudý počet $2m$ dílků, kde $m \geq 1$. Označíme $h = (b - a)/2m$ a $x_i = a + ih$, $i = 0, \dots, 2m$. Nyní na každém intervalu $\langle x_{2i}, x_{2i+2} \rangle$, $i = 0$ až $m - 1$, použijeme obdélníkové pravidlo. Dostaneme:

$$\int_{x_{2i}}^{x_{2i+2}} f(x) dx = 2hy_{2i+1} + \frac{1}{3} h^3 f''(\xi_i), \quad x_{2i} < \xi_i < x_{2i+2}, \quad i = 0, \dots, m - 1.$$

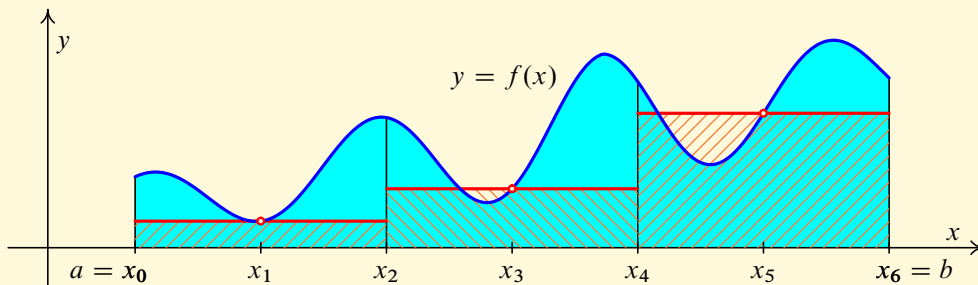
Sečtením obdržíme *složené obdélníkové pravidlo*

$$I(f) = 2h(y_1 + y_3 + \dots + y_{2m-1}) + R(f). \quad (5.15)$$

O zbytku lze snadno ukázat (viz [13, str. 70]), že ho lze vyjádřit ve tvaru

$$R(f) = \frac{b-a}{6} h^2 f''(\xi), \quad \text{kde } a < \xi < b. \quad (5.16)$$

Pravidlo je pro $2m = 6$ znázorněno na obr. 5.7.



Obr. 5.7: Složené obdélkové pravidlo

Složené lichoběžníkové pravidlo

Interval $\langle a, b \rangle$ rozdělíme na m dílků, kde $m \geq 1$. Označíme $h = (b - a)/m$ a $x_i = a + ih$, $i = 0, \dots, m$. Nyní na každém intervalu $\langle x_i, x_{i+1} \rangle$, $i = 0, \dots, m - 1$, použijeme lichoběžníkové pravidlo (5.6). Dostaneme:

$$\int_{x_i}^{x_{i+1}} f(x) dx = \frac{h}{2} (y_i + y_{i+1}) - \frac{1}{12} h^3 f''(\xi_i), \quad x_i < \xi_i < x_{i+1}, \quad i = 0, \dots, m - 1.$$

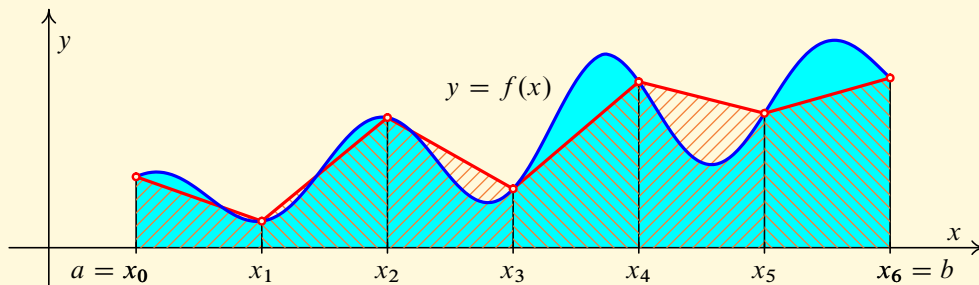
Sečtením obdržíme *složené lichoběžníkové pravidlo*

$$I(f) = \frac{h}{2} (y_0 + 2y_1 + \dots + 2y_{m-1} + y_m) + R(f). \quad (5.17)$$

O zbytku lze snadno ukázat, že ho lze vyjádřit ve tvaru

$$R(f) = -\frac{b-a}{12} h^2 f''(\xi), \quad \text{kde } a < \xi < b. \quad (5.18)$$

Pravidlo je pro $m = 6$ znázorněno na obr. 5.8.



Obr. 5.8: Složené lichoběžníkové pravidlo

Složené Simpsonovo pravidlo

Protože Simpsonovo pravidlo (5.7) dělí základní interval na dvě části, rozdělíme celý interval $\langle a, b \rangle$ na sudý počet $2m$ dílků, kde $m \geq 1$. Označíme $h = (b - a)/2m$ a $x_i = a + ih$, $i = 0, \dots, 2m$. Nyní na každém intervalu $\langle x_{2i}, x_{2i+2} \rangle$, $i = 0$ až $m - 1$, použijeme Simpsonovo pravidlo. Dostaneme:

$$\int_{x_{2i}}^{x_{2i+2}} f(x) dx = \frac{h}{3} (y_{2i} + 4y_{2i+1} + y_{2i+2}) - \frac{1}{90} h^5 f^{(4)}(\xi_i),$$

$$x_{2i} < \xi_i < x_{2i+2}, \quad i = 0, \dots, m - 1.$$

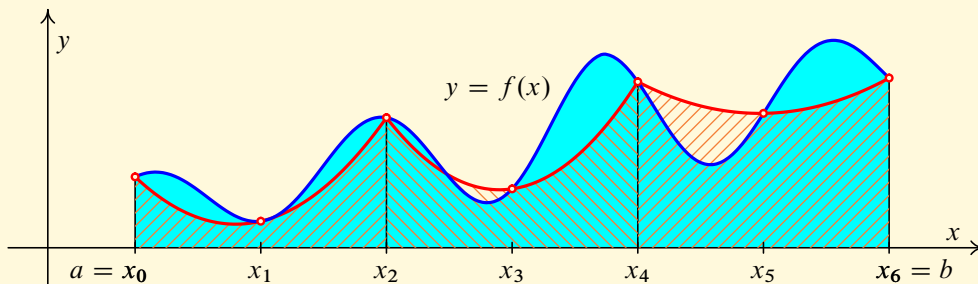
Sečtením obdržíme *složené Simpsonovo pravidlo*

$$\begin{aligned} I(f) &= \frac{h}{3} (y_0 + 4y_1 + 2y_2 + \dots + 2y_{2m-2} + 4y_{2m-1} + y_{2m}) + R(f) = \\ &= \frac{h}{3} [y_0 + y_{2m} + 4(y_1 + y_3 + \dots + y_{2m-1}) + \\ &\quad + 2(y_2 + y_4 + \dots + y_{2m-2})] + R(f). \end{aligned} \tag{5.19}$$

O zbytku lze snadno ukázat (viz [13, str. 72]), že ho lze vyjádřit ve tvaru

$$R(f) = -\frac{b-a}{180} h^4 f^{(4)}(\xi), \quad \text{kde } a < \xi < b. \quad (5.20)$$

Pravidlo je pro $2m = 6$ znázorněno na obr. 5.9.



Obr. 5.9: Složené Simpsonovo pravidlo

Vzorce pro chybu v uvedených složených formulích umožňují odhadnout, jak zvolit číslo m , abychom dosáhli potřebné přesnosti, pokud umíme odhadnout velikost derivací figurujících v těchto vzorcích. Označme

$$M_2 = \max_{x \in \langle a, b \rangle} |f''(x)|, \quad M_4 = \max_{x \in \langle a, b \rangle} |f^{(4)}(x)|.$$

Ze vzorců (5.16), (5.18) a (5.20) dostaneme

$$|R(f)| \leq \frac{b-a}{6} M_2 h^2 \quad (\text{obdélníková formule}),$$

$$|R(f)| \leq \frac{b-a}{12} M_2 h^2 \quad (\text{lichoběžníková formule}),$$

$$|R(f)| \leq \frac{b-a}{180} M_4 h^4 \quad (\text{Simpsonova formule}).$$

Dosadíme-li do těchto nerovností $h = (b-a)/m$ (lichoběžníková formule) resp. $h = (b-a)/2m$ (obdélníková a Simpsonova formule), je snadné určit m tak, aby platilo $|R(f)| < \varepsilon$, kde $\varepsilon > 0$ je předem dané malé číslo. Odhad bude obvykle pesimistický a číslo m zbytečně velké. Je to důsledek toho, že při odhadu velikosti derivace v neznámém čísle ξ jsme počítali s nejhorší (největší) možnou hodnotou.

5.4 Gaussovy kvadrurní formule

Vrátíme se k obecné kvadrurní formuli

$$Q(f) = A_0 f(x_0) + A_1 f(x_1) + \cdots + A_n f(x_n) \quad (5.21)$$

pro výpočet integrálu $\int_a^b f(x) dx$. V oddílu 5.3.1, věnovaném uzavřeným Newtonovým-Cotesovým formulím, jsme viděli, že když zvolíme ekvidistantní uzly x_i a koeficienty A_i najdeme integraci interpolačního polynomu určeného body $[x_i, f(x_i)]$, dostaneme kvadrurní formule, jejichž stupeň přesnosti je n pro lichá n a $n + 1$ pro sudá n . Otázkou je, zda je možné vybrat uzly a koeficienty nějak jinak, aby stupeň přesnosti byl vyšší. Podle věty 5.2 víme, že to nemůže být více než $2n + 1$. Není však jasné, zda lze této hodnoty dosáhnout. Kvadrurní formule dosahující tento maximální stupeň přesnosti se nazývají *Gaussovy*.

obecných mezí a a b se totiž převede na tento případ lineární substitucí:

$$\int_a^b f(x) dx = \left| \begin{array}{l} x = a + \frac{b-a}{2} (t + 1) \\ dx = \frac{b-a}{2} dt \\ a \rightsquigarrow -1, b \rightsquigarrow 1 \end{array} \right| = \frac{b-a}{2} \int_{-1}^1 f\left[a + \frac{b-a}{2} (t + 1)\right] dt.$$

Legendrový polynomy

Polynomy definované vztahy $P_0(x) = 1$ a

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n, \quad n = 1, 2, \dots, \quad (5.22)$$

se nazývají *Legendrový¹ polynomy*. Jak uvidíme dále, je možné je definovat i jinými způsoby.

¹ **Adrien-Marie Legendre** (1752–1833) (čti ležandr) — francouzský matematik. Zabýval se teorií čísel a eliptickými funkcemi.

Ze vzorce (5.22) snadno najdeme prvních šest Legendrových polynomů:

$$\begin{aligned} P_0(x) &= 1, & P_3(x) &= \frac{5}{2}x^3 - \frac{3}{2}x, \\ P_1(x) &= x, & P_4(x) &= \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8}, \\ P_2(x) &= \frac{3}{2}x^2 - \frac{1}{2}, & P_5(x) &= \frac{63}{8}x^5 - \frac{35}{4}x^3 + \frac{15}{8}x. \end{aligned} \quad (5.23)$$

Legendrové polynomy mají následující vlastnosti (viz např. [22, str. 218], [52, str. 172]) nebo [53].

- 1) Stupeň $P_n(x)$ je n . Koeficient u nejvyšší mocniny je $\frac{(2n)!}{2^n(n!)^2}$.
- 2) Platí rekurentní vzorec

$$P_{n+1}(x) = \frac{2n+1}{n+1}xP_n(x) - \frac{n}{n+1}P_{n-1}(x), \quad n = 1, 2, \dots,$$

kde $P_0(x) = 1$, $P_1(x) = x$. Tento vztah se často používá jako definice Legendrových polynomů.

- 3) Z předchozího vztahu je snadno vidět, že pro sudé n obsahuje $P_n(x)$ jen sudé mocniny x a pro liché n jen liché mocniny x .

4) $P_n(x)$ vyhovuje homogenní lineární diferenciální rovnici druhého řádu

$$(1 - x^2)y'' - 2xy' + n(n + 1)y = 0.$$

5) Všechny kořeny $P_n(x)$ jsou *reálné a jednoduché* a leží v otevřeném intervalu $(-1, 1)$. Jsou *souměrné* vzhledem k počátku a pro *liché* n je *prostřední kořen nula*.

6) Jsou-li $\bar{x}_0 < \bar{x}_1 < \dots < \bar{x}_{n-1}$ kořeny $P_n(x)$ a $x_0 < x_1 < \dots < x_{n-1} < x_n$ kořeny $P_{n+1}(x)$, $n \geq 1$, platí $x_0 < \bar{x}_0 < x_1 < \bar{x}_1 < \dots < x_{n-1} < \bar{x}_{n-1} < x_n$. Tedy kořeny $P_n(x)$ a $P_{n+1}(x)$ se pravidelně střídají.

7) Pro $m \neq n$ platí

$$\int_{-1}^1 P_m(x) P_n(x) dx = 0.$$

Říkáme, že Legendrovy polynomy tvoří *ortogonální systém*.

8) Platí

$$\int_{-1}^1 P_n^2(x) dx = \frac{2}{2n + 1}.$$

9) Je-li $Q_n(x)$, $n = 0, 1, 2, \dots$, posloupnost polynomů taková, že stupeň $Q_n(x)$ je n a pro $m \neq n$ platí

$$\int_{-1}^1 Q_m(x) Q_n(x) dx = 0,$$

splývají tyto polynomy až na konstantní násobky s Legendrovými polynomy. Existují tedy nenulové konstanty α_n takové, že $P_n(x) = \alpha_n Q_n(x)$. Vynásobíme-li tedy $Q_n(x)$ vhodnými konstantami tak, aby koeficienty u nejvyšších mocnin byly $\frac{(2n)!}{2^n (n!)^2}$, bude platit $P_n(x) = Q_n(x)$. Tuto vlastnost lze tudíž rovněž použít jako definici Legendrových polynomů.

Nyní již můžeme dát odpověď na otázku, jak zvolit uzly a koeficienty kvadraturní formule (5.21), aby stupeň přesnosti byl nejvyšší možný (viz [22, str. 231]).

Věta 5.7 *Nechť kvadraturní formule (5.21) pro výpočet integrálu $\int_{-1}^1 f(x) dx$ má stupeň přesnosti alespoň n . Pak tato formule má stupeň přesnosti $2n + 1$ právě tehdy, když její uzly jsou kořeny Legendrova polynomu $P_{n+1}(x)$.*

Je tedy jasné, že za uzly hledané kvadraturní formule s maximálním možným stupněm přesnosti je třeba zvolit kořeny polynomu $P_{n+1}(x)$. Jak víme, tento polynom má $n + 1$ různých reálných kořenů, které leží v intervalu $(-1, 1)$.

Zbývá ještě určit koeficienty A_i . Podle věty 5.3 kvadraturní formule, kterou dostaneme integrací interpolačního polynomu, má stupeň aspoň n . Podle věty 5.7 pak bude formule s takto získanými koeficienty mít maximální možný stupeň přesnosti $2n + 1$. Ze vztahu (5.3) pak navíc plyne, že koeficienty A_i jsou určeny jednoznačně.

Formule pro výpočet integrálu $\int_{-1}^1 f(x) dx$, které mají maximální možný stupeň přesnosti, se nazývají *Gaussovy-Legendrovy kvadraturní formule*.

Všimneme si praktického nalezení uzlů, koeficientů a chyby Gaussových-Legendrových kvadraturních formulí.

- Kořeny Legendrových polynomů $P_n(x)$ pro $n \leq 5$ lze snadno najít, výpočet vede na kvadratickou rovnici. Pro větší n je nutné užít numerické metody. Označme $x_0 < x_1 < \dots < x_n$ kořeny polynomu $P_{n+1}(x)$, $n \geq 0$.
- Koeficienty A_i lze získat integrací Lagrangeova polynomu. Existují však lepší možnosti. Platí (viz [22, str. 235] a [43, str. 118 a 172]), že

$$A_i = \frac{2(1 - x_i^2)}{(n + 1)^2 [P_n(x_i)]^2}, \quad i = 0, 1, \dots, n. \quad (5.24)$$

- Pro chybu lze odvodit následující vyjádření (viz [22, str. 236]):

$$R(f) = \frac{2^{2n+3} [(n + 1)!]^4}{(2n + 3) [(2n + 2)!]^2} f^{(2n+2)}(\xi), \quad \text{kde } \xi \in (-1, 1). \quad (5.25)$$

Poznámka 5.8 Ze vzorce (5.24) je vidět, že koeficienty A_i jsou kladné. To je důležité, protože tato vlastnost zaručuje, že výpočet integrálu pomocí takové kvadraturní formule je dobře podmíněná úloha (viz str. 550 nebo [13, str. 76]).

Určíme kořeny a koeficienty Gaussových-Legendrových formulí pro $n = 0, \dots, 4$. K tomu je třeba určit kořeny polynomů $P_1(x), \dots, P_5(x)$ uvedených v (5.23). Dále pak pomocí vzorce (5.24) vypočítáme koeficienty A_i . Dostaneme:

$$n = 0 \quad x_0 = 0,$$

$$A_0 = 2,$$

$$n = 1 \quad x_1 = -x_0 = \frac{1}{3} \sqrt{3} \doteq 0,577\,350,$$

$$A_0 = A_1 = 1,$$

$$n = 2 \quad x_1 = 0, \quad x_2 = -x_0 = \frac{1}{5} \sqrt{15} \doteq 0,774\,597,$$

$$A_1 = \frac{8}{9}, \quad A_0 = A_2 = \frac{5}{9},$$

$$n = 3 \quad x_3 = -x_0 = \frac{1}{35} \sqrt{525 + 70\sqrt{30}} \doteq 0,861\,136,$$

$$x_2 = -x_1 = \frac{1}{35} \sqrt{525 - 70\sqrt{30}} \doteq 0,339\,981,$$

$$A_0 = A_3 = \frac{1}{2} - \frac{1}{36} \sqrt{30} \doteq 0,347\,855,$$

$$A_1 = A_2 = \frac{1}{2} + \frac{1}{36} \sqrt{30} \doteq 0,652\,145,$$

$$\begin{aligned}
 n = 4 \quad x_2 = 0, \quad x_4 = -x_0 &= \frac{1}{21} \sqrt{245 + 14\sqrt{70}} \doteq 0,906\,180, \\
 x_3 = -x_1 &= \frac{1}{21} \sqrt{245 - 14\sqrt{70}} \doteq 0,538\,469, \\
 A_2 = \frac{128}{225}, \quad A_0 = A_4 &= \frac{161}{450} - \frac{13}{900} \sqrt{70} \doteq 0,236\,927, \\
 A_1 = A_3 &= \frac{161}{450} + \frac{13}{900} \sqrt{70} \doteq 0,478\,629.
 \end{aligned}$$

Nakonec napíšeme Gaussovy-Legendrovy kvadraturní formule pro $n = 0, 1, 2$ včetně chyby, kterou určíme ze vztahu (5.25). Ve všech vzorcích je $-1 < \xi < 1$.

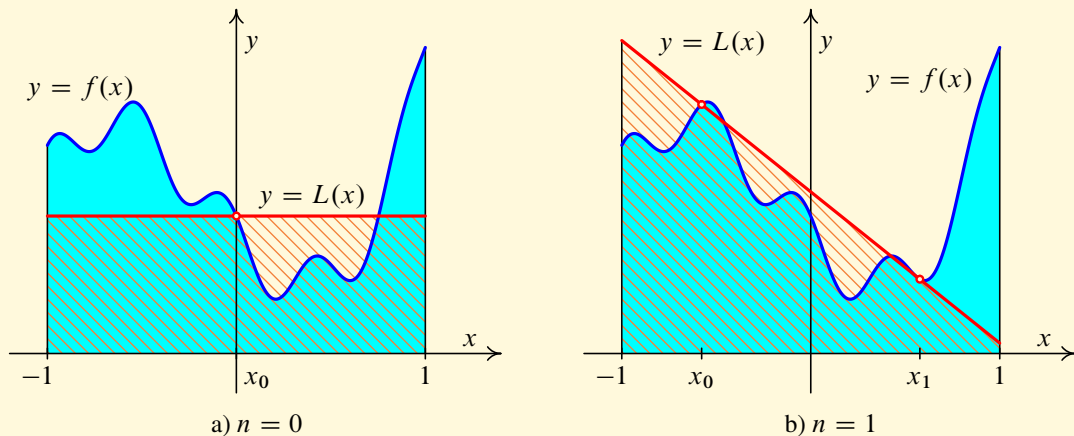
$$I(f) = 2f(0) + \frac{1}{3} f''(\xi), \quad (5.26)$$

$$I(f) = f\left(-\frac{1}{3}\sqrt{3}\right) + f\left(\frac{1}{3}\sqrt{3}\right) + \frac{1}{135} f^{(4)}(\xi), \quad (5.27)$$

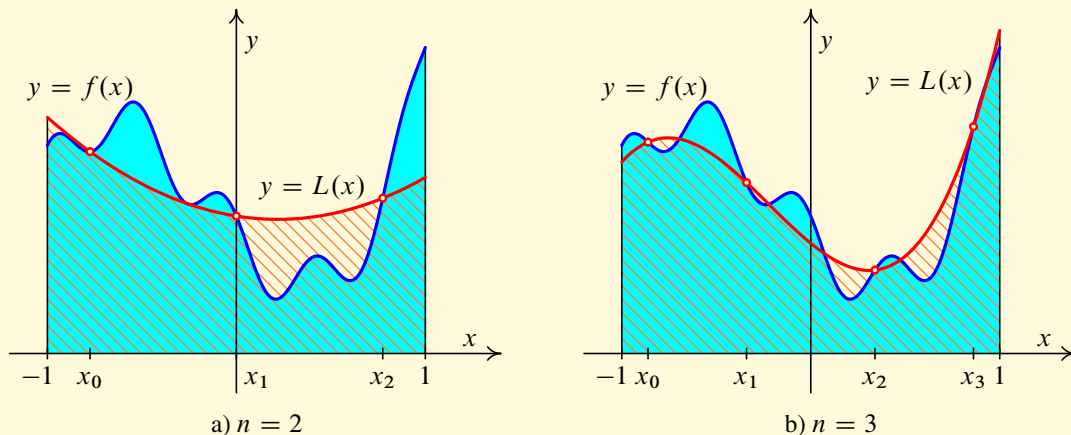
$$I(f) = \frac{5}{9} f\left(-\frac{1}{5}\sqrt{15}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\frac{1}{5}\sqrt{15}\right) + \frac{1}{15750} f^{(6)}(\xi). \quad (5.28)$$

Obdobně by bylo možné zapsat tyto formule pro vyšší n .

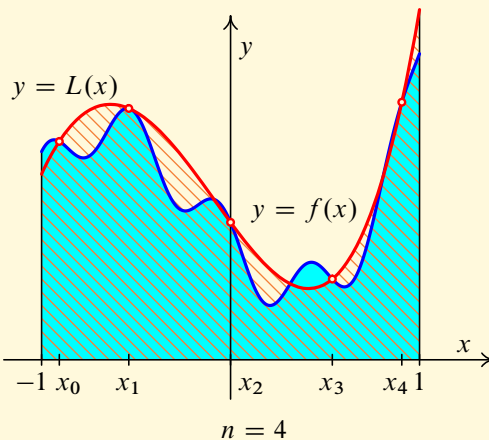
Geometrický význam Gaussových-Legendrových formulí pro $n = 0, \dots, 4$ je znázorněn na obr. 5.10, 5.11 a 5.12. Z obrázku 5.10 a) i ze vzorce (5.26) je zřejmé, že pro $n = 0$ jde vlastně o obdélníkové pravidlo, srovnejte s obr. 5.4 a) a vzorcem (5.11).



Obr. 5.10: Gaussovy-Legendrovy kvadraturní formule — část 1



Obr. 5.11: Gaussovy-Legendrovy kvadraturní formule — část 2



Obr. 5.12: Gaussovy-Legendrovy kvadraturní formule — část 3

Poznámka 5.9 V poznámce 5.5 jsme se zmínili o výpočtu integrálů typu $\int_a^b w(x) f(x) dx$ s váhou $w(x)$. I pro tyto integrály je možné hledat Gaussovy formule, tj. kvadraturní formule s maximálním možným stupněm přesnosti. Obdobný je i výsledek, za uzly je nutné volit kořeny jistých speciálních polynomů. Zejména se užívají tyto případy:

- $w(x) = 1/\sqrt{1-x^2}$, $a = -1$, $b = 1$ — Gaussovy-Čebyševovy¹ formule,
- $w(x) = e^{-x}$, $a = 0$, $b = \infty$ — Gaussovy-Laguerrovy² formule,
- $w(x) = e^{-x^2}$, $a = -\infty$, $b = \infty$ — Gaussovy-Hermitovy³ formule.

Podrobněji viz např. [22, str. 231] nebo [52, str. 171].

¹**Pafnutij Lvovič Čebyšev** (1821–1894) — významný ruský matematik. Zabýval se analýzou, teorií čísel a pravděpodobností.

²**Edmond Nicolas Laguerre** (1834–1886) (čti lager) — francouzský matematik. Zabýval se geometrií a komplexní analýzou.

³Viz str. 338.

5.5 Rombergova kvadratura

Metoda, kterou popíšeme, vychází se složeného lichoběžníkového pravidla (5.17). Protože budeme pracovat s různým počtem dílků, označíme

$$Q_h(f) = \frac{h}{2} (y_0 + 2y_1 + \cdots + 2y_{m-1} + y_m),$$

kde $h = (b - a)/m$.

Lze dokázat (viz [52, str. 160], důkaz je poměrně obtížný), že pro funkci $f(x)$, která má spojité derivace na intervalu $\langle a, b \rangle$ až do řádu $2k + 2$, $k \geq 1$, je možné kvadrurní formuli $Q_h(f)$ vyjádřit ve tvaru

$$Q_h(f) = c_0 + c_1 h^2 + \cdots + c_k h^{2k} + \alpha_{k+1}(h) h^{2k+2}, \quad (5.29)$$

kde c_0, \dots, c_k jsou konstanty (závisející na a, b a $f(x)$). Přitom

$$c_0 = \int_a^b f(x) dx = I(f).$$

Funkce $\alpha_{k+1}(h)$ (závisející na a , b a $f(x)$) je ohraničená konstantou M_{k+1} , tudíž platí $|\alpha_{k+1}(h)| \leq M_{k+1}$, přičemž tato konstanta je *univerzální* pro libovolné h , nezávisí tedy na počtu dílků m . Vztah (5.29) se nazývá *asymptotickým rozvojem formule $Q_h(f)$ v proměnné h* .

Jsou známy explicitní vzorce pro c_1, \dots, c_k a $\alpha_{k+1}(h)$. Jsou však dost komplikované, a protože je nebudeme potřebovat, nebudeme je uvádět.

Označme

$$P_k(y) = c_0 + c_1 y + \dots + c_k y^k.$$

Pak $P_k(y)$ je polynom stupně nejvýše k a platí $P_k(0) = c_0 = \int_a^b f(x) dx$. Pokusíme se přibližně najít hodnotu $P_k(0)$.

Pro malá h je chybový člen $\alpha_{k+1}(h)h^{2k+2}$ v (5.29) malý, takže platí přibližný vztah

$$Q_h(f) \approx P_k(h^2).$$

Nyní zvolíme posloupnost kroků $h_0 = b - a$, $h_1 = (b - a)/2$, \dots , $h_k = (b - a)/2^k$. Tedy následující krok má ve srovnání s předchozím poloviční délku. Vypočítáme složené lichoběžníkové pravidlo pro tyto hodnoty. Dostaneme:

$$Q_{h_i}(f) \approx P_k(h_i^2), \quad i = 0, \dots, k.$$

Předpokládejme, že sestrojíme interpolační polynom určený dvojicemi $[y_i, P(y_i)]$, kde $y_i = h_i^2$. Jeho tvar podle (4.3) bude

$$L(x) = P(y_0)L_0(x) + P(y_1)L_1(x) + \cdots + P(y_k)L_k(x),$$

kde $L_i(x)$ jsou speciální interpolační polynomy. Protože stupeň $L(x)$ je nejvýše k , plyne z jednoznačnosti interpolačního polynomu (věta 4.1), že $P_k(x) = L(x)$, tj.

$$P_k(x) = P(y_0)L_0(x) + P(y_1)L_1(x) + \cdots + P(y_k)L_k(x).$$

Po dosazení $y_i = h_i^2$ pro $x = 0$ dostaneme

$$P_k(0) = P(h_0^2)L_0(0) + P(h_1^2)L_1(0) + \cdots + P(h_k^2)L_k(0).$$

Přesné hodnoty $P(h_i^2)$ však neznáme. Nahradíme je tudíž přibližnými hodnotami $Q_{h_i}(f)$. Pro výpočet čísla $P_k(0)$ tak máme přibližný vzorec

$$P_k(0) \approx Q_{h_0}(f)L_0(0) + Q_{h_1}(f)L_1(0) + \cdots + Q_{h_k}(f)L_k(0).$$

Označíme-li $d_i = L_i(0)$, $i = 0, \dots, k$, dostáváme pro výpočet hledaného integrálu formuli

$$\int_a^b f(x) dx \approx d_0 Q_{h_0}(f) + d_1 Q_{h_1}(f) + \dots + d_k Q_{h_k}(f), \quad (5.30)$$

který se nazývá *Rombergova*¹ *kvadraturní formule*.

Abychom formuli mohli použít, měli bychom najít interpolační polynom (Lagrangeovým nebo Newtonovým algoritmem) a dosadit do něj nulu. To je však zbytečně pracné. Vzhledem k tomu, že potřebujeme najít funkční hodnotu v jediném bodě, je vhodnější použít Nevillův algoritmus pro iterovanou interpolaci (viz str. 317). Protože ten jsme neprobírali, popíšeme nyní, jak na něm založený výpočet v případě Rombergovy kvadratury probíhá.

Označíme $T_{i0} = Q_{h_i}(f)$, $i = 0, \dots, k$. Tedy T_{i0} je přibližná hodnota hledaného integrálu získaná složeným lichoběžníkovým pravidlem, kde základní interval $\langle a, b \rangle$ jsme rozdělili na 2^i dílků. (Písmeno T se používá, protože lichoběžník je anglicky trapezium.)

¹**Werner Romberg** (1909–2003) — německý matematik. Zabýval se numerickou matematikou. Vzorec navrhl v roce 1955.

Dále vypočítáme následující čísla:

$$T_{ij} = \frac{4^j T_{i,j-1} - T_{i-i,j-1}}{4^j - 1}, \quad i = 1, \dots, k, \quad j = 1, \dots, i. \quad (5.31)$$

Pak platí, že

$$T_{kk} = d_0 Q_{h_0}(f) + d_1 Q_{h_1}(f) + \dots + d_k Q_{h_k}(f),$$

tedy

$$\int_a^b f(x) dx \approx T_{kk}.$$

Hodnoty T_{ij} obvykle zapisujeme do trojúhelníkového schématu

$$\begin{array}{cccccc} T_{00} & & & & & \\ T_{10} & T_{11} & & & & \\ T_{20} & T_{21} & T_{22} & & & \\ T_{30} & T_{31} & T_{32} & T_{33} & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ T_{k0} & T_{k1} & T_{k2} & T_{k3} & \dots & T_{kk} \end{array}$$

Nejprve vypočítáme složeným lichoběžníkovým pravidlem hodnoty v levém sloupci. Zbytek tabulky pak vyplňujeme po řádcích užitím vzorce (5.31). Následující schéma ukazuje, které dvě hodnoty se používají při výpočtu T_{ij} .

$$\begin{array}{ccc} T_{i-1,j-1} & & \\ & \searrow & \\ T_{i,j-1} & \rightarrow & T_{ij} \end{array}$$

Nevillův algoritmus umožňuje snadno přidat další hodnotu, Stačí vypočítat složeným lichoběžníkovým pravidlem číslo $T_{k+1,0}$ a pomocí vzorce (5.31) dopočítat další řádek. Jako zastavovací kritérium lze použít například podmínku $|T_{kk} - T_{k-1,k-1}| < \varepsilon$ nebo $|T_{kk} - T_{k,k-1}| < \varepsilon$, kde $\varepsilon > 0$ je předepsaná tolerance.

Rombergova formule se používá, vyžadujeme-li vysokou přesnost. Je však třeba mít na paměti, že je založena na vzorci (5.29), jehož předpokladem je dostatečná hladkost funkce $f(x)$. Není tedy vhodná pro funkce, které tento předpoklad nesplňují. V takovém případě nepřináší žádné zlepšení oproti složené lichoběžníkové metodě.

Rombergova metoda patří mezi tzv. *extrapolační metody*. Pomocí interpolačního

polynomu, který je určen hodnotami v kladných uzlových bodech

$$h_k^2 < h_{k-1}^2 < \cdots < h_1^2 < h_0^2,$$

totiž počítá hodnotu polynomu $P_k(y)$ v čísle nula, které je mimo interval $\langle h_k^2, h_0^2 \rangle$. Další informace o této metodě lze nalézt např. v [22, str. 266] nebo [52, str. 160].

Poznámka 5.10 Rombergova kvadratura vychází z tzv. *Richardsonovy*¹ *extrapolace*. Jedná se o metodu, která je v numerické matematice často používaná. Slouží k tomu, aby pro veličinu, která má asymptotický rozvoj podobného typu, jako je (5.29), a jejíž hodnoty na diskrétní množině kladných čísel umíme určit, bylo možné vytvořit formule, které dávají hodnotu v nule s vyšším řádem přesnosti. Podrobněji viz [22, str. 212] nebo [52, str. 165].

¹**Lewis Fry Richardson** (1881–1953) (čti ričrdson) — anglický matematik, fyzik a meteorolog.

Příklad 5.11 Pomocí Rombergovy kvadraturní formule vypočítejte hodnotu integrálu $\int_0^2 x e^{-x^2} \sin(2x^2) dx$. Volte $k = 4, 5, 6$.

Řešení. Integrand má derivace všech řádů, takže za k lze volit libovolně velké číslo. Výpočet byl proveden na počítači, mantisa byla nastavena na 8.

$$T_{00} = 0,0362415$$

$$T_{10} = 0,3526326 \quad T_{11} = 0,4580963$$

$$T_{20} = 0,1923873 \quad T_{21} = 0,1389723 \quad T_{22} = 0,1176973$$

$$T_{30} = 0,1977703 \quad T_{31} = 0,1995647 \quad T_{32} = 0,2036042 \quad T_{33} = 0,2049678$$

$$T_{40} = 0,1984967 \quad T_{41} = 0,1987388 \quad T_{42} = 0,1986837 \quad T_{43} = 0,1986056 \quad T_{44} = 0,1985807$$

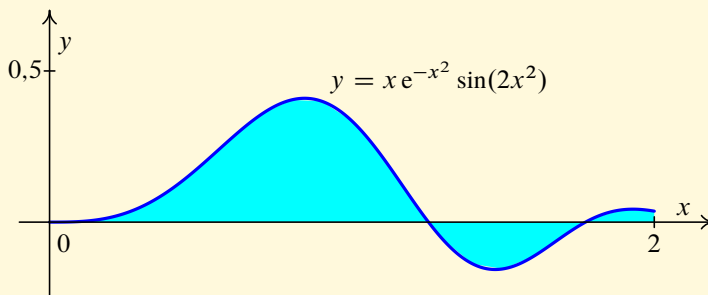
Integrál lze elementárními metodami vypočítat, přesná hodnota je

$$\frac{1}{5} [1 - e^{-4}(2 \cos^2 4 + \cos 4 \sin 4 - 1)] \doteq 0,198\,720\,912\,4.$$

Analogicky pro $k = 5$ vyjde 0,198 721 9 a pro $k = 6$ vyjde 0,198 720 9, takže v tomto případě už jsou všechny cifry platné.

Graf integrandu je znázorněn na obr. 5.13





Obr. 5.13: Graf funkce z příkladu 5.11

5.6 Numerická derivace

Nutnost numerického výpočtu derivace a nemožnost použití standardních vzorců pro derivace elementárních funkcí a pravidel pro derivování má obdobné příčiny jako u integrálu. Důvody jsou zejména následující:

- (1) Ve vzorci mohou být obsaženy funkce, jejichž derivace neumíme vyjádřit pomocí elementárních funkcí. Například to mohou být Eulerovy funkce gama a beta, řešení různých diferenciálních rovnic (třeba Besselovy funkce) atd.
- (2) Vzorec funkce $f(x)$ sice máme, ale výpočet derivace standardními postupy by byl velmi pracný.
- (3) Vzorec funkce $f(x)$ nemáme a její hodnoty známe pouze na konečné množině bodů (např. jsou to výsledky nějakého měření nebo algoritmu).

Předpokládejme, že známe hodnoty $f(x_0), f(x_1), \dots, f(x_n)$ v uzlových bodech x_0, x_1, \dots, x_n , které jsou po dvou různé. Podle vzorce (4.3) sestrojíme Lagrangeův inter-

polační polynom

$$L(x) = y_0 L_0(x) + \cdots + y_n L_n(x).$$

Ten použijeme k přibližnému určení derivací funkce $f(x)$ na intervalu $\text{Int}\langle x_0, \dots, x_n \rangle$ (následující úvahy platí i mimo tento interval, ale tam interpolační polynom obecně není dobrou aproximací funkce $f(x)$). Hodnotu k -té derivace $f^{(k)}(x)$ nahradíme hodnotou $L^{(k)}(x)$.

Bude nás bude zajímat chyba, které se při této náhradě dopustíme. Jak jsme si na začátku této kapitoly ukázali, z toho, že dvě funkce mají blízké hodnoty, neplyne, že jsou blízké i hodnoty jejich derivací. Navíc víme, že $f(x) \approx L(x)$ nemusí platit na celém intervalu $\text{Int}\langle x_0, \dots, x_n \rangle$, pokud je počet uzlů $n + 1$ velký (viz příklad 4.5).

Podle věty 4.4 pro chybu aproximace interpolačním polynomem platí vztah

$$f(x) = L(x) + \omega(x) \frac{f^{(n+1)}(\xi)}{(n+1)!},$$

kde $\omega(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$ a $\xi \in \text{Int}\langle x_0, \dots, x_n, x \rangle$ je nějaké číslo závislé na x a funkce $f(x)$ má $(n + 1)$ -ní derivaci. Chyba aproximace $R(x)$ v bodě x

je tudíž dána vzorcem

$$R(x) = \omega(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}. \quad (5.32)$$

Předpokládejme nyní, že funkce $f(x)$ má na nějakém intervalu J obsahujícím všechny uzlové body x_i , $i = 0, \dots, n$, derivace až do řádu k (musí mít tedy derivace až do řádu $\max(n+1, k)$). Jelikož $L(x)$ je polynom a platí rovnost $R(x) = f(x) - L(x)$, má také chyba $R(x)$ na J derivace až do řádu k . Platí tedy, že

$$f^{(j)}(x) = L^{(j)}(x) + \frac{d^j}{dx^j} \left[\omega(x) \frac{f^{(n+1)}(\xi)}{(n+1)!} \right], \quad j = 1, \dots, k, \quad x \in J. \quad (5.33)$$

Chyba $R(x)$ má tvar součinu. První činitel je mnohočlen $\omega(x)$, který má derivace libovolného řádu. Druhý činitel je však až na konstantu $1/(n+1)!$ výraz $f^{(n+1)}(\xi)$, v němž ξ je funkce závisující na x , o níž nemáme žádné informace. Není tedy jasné, zda má tento činitel derivaci, a pokud ano, jak by se měla vypočítat (vzorec pro derivaci složené funkce nelze použít). Nicméně je možné dokázat následující výsledek.

Věta 5.12 Má-li funkce $f(x)$ má na intervalu J obsahujícím všechny uzlové body x_i , $i = 0, \dots, n$, derivaci řádu $n + k + 1$, pak při označení z (5.33) platí, že

$$\frac{d^j}{dx^j} \left[\frac{f^{(n+1)}(\xi)}{(n+1)!} \right] = \frac{j!}{(n+1+j)!} f^{(n+1+j)}(\eta_j), \quad j = 1, \dots, k, x \in J,$$

kde $\eta_j \in \text{Int}\langle x_0, \dots, x_n, x \rangle$ jsou čísla závisající na x . Je-li tento interval nedegenerovaný, je η_j jeho vnitřní bod (k degeneraci intervalu dojde pro $n = 0$ a $x = x_0$, pak $\eta_j = x_0$).

Důkaz tvrzení je technicky poměrně náročný, viz [3, str. 46 a 89], [43, str. 100], pro $k = 1$ také [22, str. 206].

S pomocí předchozího výsledku nyní můžeme spočítat derivace chyby $R(x)$. Výsledky vypíšeme pro první a druhou derivaci podrobně. Z (5.32) dostaneme:

$$R'(x) = \omega'(x) \frac{f^{(n+1)}(\xi_0)}{(n+1)!} + \omega(x) \frac{f^{(n+2)}(\xi_1)}{(n+2)!}, \quad (5.34)$$

$$R''(x) = \omega''(x) \frac{f^{(n+1)}(\eta_0)}{(n+1)!} + 2\omega'(x) \frac{f^{(n+2)}(\eta_1)}{(n+2)!} + \omega(x) \frac{2f^{(n+3)}(\eta_2)}{(n+3)!}, \quad (5.35)$$

kde $\xi_0, \xi_1, \eta_0, \eta_1, \eta_2 \in \text{Int}\langle x_0, \dots, x_n, x \rangle$.

S použitím Leibnizova vzorce pro derivaci součinu (viz např. [16, str. 112]) obecně dostaneme pro $j = 1, \dots, k$ a $x \in J$:

$$\begin{aligned} R^{(j)}(x) &= \sum_{i=0}^j \binom{j}{i} \omega^{(j-i)}(x) \frac{i!}{(n+1+i)!} f^{(n+1+i)}(\eta_i) = \\ &= \sum_{i=0}^j \frac{j!}{(j-i)!(n+1+i)!} \omega^{(j-i)}(x) f^{(n+1+i)}(\eta_i), \end{aligned} \quad (5.36)$$

kde $\eta_0, \dots, \eta_j \in \text{Int}\langle x_0, \dots, x_n, x \rangle$.

Pro $x \notin \text{Int}\langle x_0, \dots, x_n \rangle$, tj. v případě extrapolace, lze předchozí výsledek zjednodušit (x může být některým z krajních bodů tohoto intervalu). Lze dokázat následující větu.

Věta 5.13 Má-li funkce $f(x)$ má na intervalu J obsahujícím všechny uzlové body x_i , $i = 0, \dots, n$, derivaci řádu $n + 1$, pak pro $x \notin \text{Int}(x_0, \dots, x_n)$ a $j = 1, \dots, n + 1$ platí

$$R^{(j)}(x) = \frac{1}{(n+1)!} \omega^{(j)}(x) f^{(n+1)}(\eta), \quad \eta \in \text{Int}\langle x_0, \dots, x_n, x \rangle, \quad (5.37)$$

přičemž pro $j < n + 1$ je η je vnitřní bod, zatímco pro $j = n + 1$ je $\eta = x$.

Návod k důkazu viz [43, str. 165].

Vzorec (5.37) říká, že pro $x \notin \text{Int}(x_0, \dots, x_n)$ lze za daných předpokladů ve vzorci (5.36) uvažovat jen první člen. Číslo η_0 z prvního vzorce je obecně jiné než číslo η ve druhém vzorci. Předchozí věta je zejména užitečná v případě, když x je některý z krajních bodů intervalu $\text{Int}\langle x_0, \dots, x_n \rangle$.

V dalším se budeme věnovat vzorcům pro nalezení přibližných hodnot první a druhé derivace.

Z (5.33), (5.34) a (5.35) dostaneme, že

$$f'(x) = L'(x) + \omega'(x) \frac{f^{(n+1)}(\xi_0)}{(n+1)!} + \omega(x) \frac{f^{(n+2)}(\xi_1)}{(n+2)!}, \quad (5.38)$$

$$f''(x) = L''(x) + \omega''(x) \frac{f^{(n+1)}(\eta_0)}{(n+1)!} + 2\omega'(x) \frac{f^{(n+2)}(\eta_1)}{(n+2)!} + \omega(x) \frac{2f^{(n+3)}(\eta_2)}{(n+3)!}. \quad (5.39)$$

Předchozí vzorce pro numerickou derivaci se nejčastěji používají pro nalezení hodnot první popř. druhé derivace v uzlových bodech x_i , $i = 0, 1, \dots, n$. Zřejmě platí, že $\omega(x_i) = 0$. Určíme, kolik je $\omega'(x_i)$. Použijeme tzv. logaritmické derivování. Platí:

$$\omega(x) = (x - x_0)(x - x_1) \cdots (x - x_n),$$

$$\ln \omega(x) = \ln(x - x_0) + \ln(x - x_1) + \cdots + \ln(x - x_n),$$

$$\frac{\omega'(x)}{\omega(x)} = \frac{1}{x - x_0} + \frac{1}{x - x_1} + \cdots + \frac{1}{x - x_n},$$

$$\omega'(x) = \omega(x) \left(\frac{1}{x - x_0} + \frac{1}{x - x_1} + \cdots + \frac{1}{x - x_n} \right) = \sum_{k=0}^n \prod_{\substack{j=0 \\ j \neq k}}^n (x - x_j).$$

V předchozím odvození je třeba, aby $x > \max\{x_0, \dots, x_n\}$, snadno se však ověří, že výsledek platí pro všechna x , jde totiž o rovnost dvou polynomů.

Pro $x = x_i$ z předcházející rovnosti vyjde (v součinu se vynechá činitel $(x - x_i)$):

$$\begin{aligned}\omega'(x_i) &= (x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n) = \\ &= \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j), \quad i = 0, 1, \dots, n.\end{aligned}\tag{5.40}$$

Vzorec (5.38) se zjednoduší na tvar

$$f'(x_i) = L'(x_i) + \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j) \cdot \frac{f^{(n+1)}(\xi)}{(n+1)!}, \quad i = 0, \dots, n,\tag{5.41}$$

kde $\xi \in \text{Int}\langle x_0, \dots, x_n \rangle$.

Dále určíme, kolik je $\omega''(x_i)$. Vzorec pro $\omega'(x)$ má tvar součtu součinů. Na každý z těchto součinů majících vždy n činitelů použijeme opět logaritmické derivování. Po

úpravách, jejichž detaily vynecháme, vyjde (v součinech vynecháme vždy dva činitele):

$$\omega''(x) = 2 \sum_{\substack{k,l=0 \\ k < l}}^n \prod_{\substack{j=0 \\ j \neq k,l}}^n (x - x_j).$$

Pro $x = x_i$ dostaneme:

$$\omega''(x_i) = 2 \sum_{\substack{k=0 \\ k \neq i}}^n \prod_{\substack{j=0 \\ j \neq i,k}}^n (x_i - x_j) = 2\omega'(x_i) \sum_{\substack{j=0 \\ j \neq i}}^n \frac{1}{x_i - x_j}, \quad i = 0, 1, \dots, n. \quad (5.42)$$

Vzorec (5.39) se zjednoduší na tvar

$$\begin{aligned} f''(x_i) &= L''(x_i) + 2 \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j) \cdot \sum_{\substack{j=0 \\ j \neq i}}^n \frac{1}{x_i - x_j} \cdot \frac{f^{(n+1)}(\xi_0)}{(n+1)!} + \\ &+ 2 \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j) \cdot \frac{f^{(n+2)}(\xi_1)}{(n+2)!}, \quad i = 0, 1, \dots, n, \end{aligned} \quad (5.43)$$

kde $\xi_0, \xi_1 \in \text{Int}\langle x_0, \dots, x_n \rangle, i = 0, \dots, n$.

Nyní odvodíme nejčastěji používané vzorce. Budeme předpokládat, že uzly jsou ekvidistantní s krokem $h > 0$, tedy $x_i = x_0 + hi, i = 0, 1, \dots, n$. V tomto případě bude výhodné použít místo Lagrangeova interpolačního polynomu $L(x)$ Newtonův interpolační polynom vpřed $N^+(t)$ (jedná se o tentýž polynom, liší se jen zápisem).

V dalším označíme pro stručnost $f_i = f(x_i)$. Podle (4.8) je

$$N^+(t) = f_0 + \binom{t}{1} \Delta f_0 + \binom{t}{2} \Delta^2 f_0 + \dots + \binom{t}{n} \Delta^n f_0, \quad (5.44)$$

kde $t = \frac{x-x_0}{h}$ a $\binom{t}{k}$ je zobecněný binomický koeficient. Pro uzlové body přitom platí $x_i - x_j = h(i - j), i, j = 0, 1, \dots, n$. Ze vzorců (5.40) a (5.42) po úpravách pro $\omega'(x_i)$ a $\omega''(x_i), i = 0, 1, \dots, n$, dostaneme vyjádření

$$\omega'(x_i) = (-1)^{n-i} i! (n - i)! h^n, \quad (5.45)$$

$$\omega''(x_i) = 2(-1)^{n-i} i! (n - i)! h^{n-1} \left(\sum_{j=1}^i \frac{1}{j} - \sum_{j=1}^{n-i} \frac{1}{j} \right). \quad (5.46)$$

S využitím vzorce pro derivaci složené funkce dále dostaneme (vnitřní složka je $t = (x - x_0)/h$):

$$L'(x) = \frac{d}{dx} N^+(t) = \frac{d}{dt} N^+(t) \frac{dt}{dx} = \frac{1}{h} \frac{d}{dt} N^+(t), \quad (5.47)$$

$$L''(x) = \frac{d^2}{dx^2} N^+(t) = \frac{1}{h} \frac{d}{dx} \frac{d}{dt} N^+(t) = \frac{1}{h} \frac{d^2}{dt^2} N^+(t) \frac{dt}{dx} = \frac{1}{h^2} \frac{d^2}{dt^2} N^+(t),$$

$$\vdots \quad (5.48)$$

$$L^{(k)}(x) = \frac{d^k}{dx^k} N^+(t) = \frac{1}{h^k} \frac{d^k}{dt^k} N^+(t). \quad (5.49)$$

Nyní ze vzorců (5.41), (5.45) a (5.47) pro první derivaci v uzlovém bodě x_i vyjde, že

$$f'(x_i) = \frac{1}{h} \frac{d}{dt} N^+(i) + (-1)^{n-i} i! (n-i)! h^n \frac{f^{(n+1)}(\xi)}{(n+1)!}, \quad (5.50)$$

kde $\xi \in \text{Int}\langle x_0, \dots, x_n \rangle$, $i = 0, \dots, n$.

Obdobně ze vzorců (5.43), (5.45), (5.46) a (5.48) pro druhou derivaci v uzlovém bodě x_i vyjde, že

$$f''(x_i) = \frac{1}{h^2} \frac{d^2}{dt^2} N^+(i) + 2(-1)^{n-i} i! (n-i)! h^n \frac{f^{(n+2)}(\xi_1)}{(n+2)!} + \\ + 2(-1)^{n-i} i! (n-i)! h^{n-1} \left(\sum_{j=1}^i \frac{1}{j} - \sum_{j=1}^{n-i} \frac{1}{j} \right) \frac{f^{(n+1)}(\xi_0)}{(n+1)!}, \quad (5.51)$$

kde $\xi_0, \xi_1 \in \text{Int}\langle x_0, \dots, x_n \rangle$, $i = 0, \dots, n$. Z předchozího vzorce je vidět, že když bude počet uzlů $n+1$ liché číslo a x_i bude prostřední uzel, bude koeficient u $\frac{f^{(n+1)}(\xi_0)}{(n+1)!}$ nulový, takže chyba bude řádu $O(h^n)$, nikoli pouze $O(h^{n-1})$.

S připravenými vzorci snadno odvodíme nejběžnější vzorce pro první a druhou derivaci využívající dva nebo tři po sobě jdoucí uzlové body. Odvození provedeme pro uzly x_0, x_1 resp. x_0, x_1, x_2 . Vzorce pro libovolnou po sobě jdoucí dvojici nebo trojici, které se z nich snadno přepíší, uvedeme nakonec v souhrnné tabulce. V dalším budeme předpokládat, že funkce $f(x)$ má derivace potřebného řádu, viz věta 5.12.

Vzorce pro numerickou derivaci používající dva uzly

Pro $n = 1$ je podle vzorce (5.44)

$$N^+(t) = f_0 + \Delta f_0 t, \quad \text{tedy} \quad \frac{d}{dt} N^+(t) = \Delta f_0 = f_1 - f_0,$$

a tudíž

$$\frac{d}{dt} N^+(0) = \frac{d}{dt} N^+(1) = f_1 - f_0.$$

Ze vzorce (5.50) pro $i = 0$ vyjde

$$f'(x_0) = \frac{1}{h} (f_1 - f_0) + (-1)^1 \cdot 0! \cdot 1! \cdot h \frac{f''(\xi)}{2!} = \frac{f_1 - f_0}{h} - \frac{h}{2} f''(\xi)$$

a pro $i = 1$ vyjde

$$f'(x_1) = \frac{1}{h} (f_1 - f_0) + (-1)^0 \cdot 1! \cdot 0! \cdot h \frac{f''(\xi)}{2!} = \frac{f_1 - f_0}{h} + \frac{h}{2} f''(\xi).$$

Vzorce pro numerickou derivaci používající tři uzly

Pro $n = 2$ je podle vzorce (5.44)

$$N^+(t) = f_0 + \Delta f_0 t + \frac{1}{2} \Delta^2 f_0 (t^2 - t),$$

tedy

$$\frac{d}{dt} N^+(t) = \Delta f_0 + \frac{1}{2} \Delta^2 f_0 (2t - 1) = f_1 - f_0 + \frac{1}{2} (f_2 - 2f_1 + f_0)(2t - 1),$$

$$\frac{d^2}{dt^2} N^+(t) = \Delta^2 f_0 = f_2 - 2f_1 + f_0,$$

a tudíž

$$\frac{d}{dt} N^+(0) = \frac{-3f_0 + 4f_1 - f_2}{2}, \quad \frac{d}{dt} N^+(2) = \frac{f_0 - 4f_1 + 3f_2}{2},$$

$$\frac{d}{dt} N^+(1) = \frac{f_2 - f_0}{2}, \quad \frac{d^2}{dt^2} N^+(0) = \frac{d^2}{dt^2} N^+(1) = \frac{d^2}{dt^2} N^+(2) = f_2 - 2f_1 + f_0.$$

Ze vzorce (5.50) pro $i = 0$ vyjde

$$\begin{aligned} f'(x_0) &= \frac{-3f_0 + 4f_1 - f_2}{2h} + (-1)^2 \cdot 0! \cdot 2! \cdot h^2 \frac{f''(\xi)}{6} = \\ &= \frac{-3f_0 + 4f_1 - f_2}{2h} + \frac{h^2}{3} f'''(\xi), \end{aligned}$$

pro $i = 1$ vyjde

$$f'(x_1) = \frac{f_2 - f_0}{2h} + (-1)^1 \cdot 1! \cdot 1! \cdot h^2 \frac{f''(\xi)}{6} = \frac{f_2 - f_0}{2h} - \frac{h^2}{6} f'''(\xi)$$

a pro $i = 2$ vyjde

$$\begin{aligned} f'(x_2) &= \frac{f_0 - 4f_1 + 3f_2}{2h} + (-1)^0 \cdot 2! \cdot 0! \cdot h^2 \frac{f''(\xi)}{6} = \\ &= \frac{f_0 - 4f_1 + 3f_2}{2h} + \frac{h^2}{3} f'''(\xi). \end{aligned}$$

a pro $i = 1$ vyjde

$$f''(x_1) = \frac{f_2 - 2f_1 + f_0}{h^2} + 2(-1)^1 \cdot 1! \cdot 1! \cdot h^2 \frac{f^{(4)}(\xi_1)}{24} + 2(-1)^1 \cdot 1! \cdot 1! \cdot h \cdot (1-1) \frac{f'''(\xi_0)}{6} = \frac{f_2 - 2f_1 + f_0}{h^2} - \frac{h^2}{12} f^{(4)}(\xi_1).$$

Výsledek v uzlu x_1 je ve shodě s komentářem za (5.51) (použili jsme tři uzly, což je lichý počet, a počítali druhou derivaci v prostředním z nich).

Všimněte si, že u dvoubodových formulí je chyba řádu $O(h)$, kdežto u třibodových je řádu $O(h^2)$ s výjimkou vzorců pro $f''(x_0)$ a $f''(x_2)$, kde je také jen $O(h)$. Tyto vzorce by bylo možné odvodit rovněž pomocí Taylorova vzorce, viz oddíl 7.2.3.

Na závěr uvedeme přehled všech osmi odvozených vzorců pro libovolné dvojice resp. trojice po sobě jdoucích uzlů. Indexaci zvolíme podle uzlu, v němž počítáme derivaci. O funkci $f(x)$ se předpokládá, že má derivace potřebného řádu. Ve vzorcích pro druhou derivaci využívajících tři uzly můžeme podle věty 5.13 vynechat u derivací v krajních uzlech ve výrazu pro chybu člen s h^2 .

$$f'(x_i) = \frac{f_{i+1} - f_i}{h} - \frac{h}{2} f''(\xi), \quad \xi \in (x_i, x_{i+1}), \quad (5.52)$$

$$f'(x_i) = \frac{f_i - f_{i-1}}{h} + \frac{h}{2} f''(\xi), \quad \xi \in (x_{i-1}, x_i), \quad (5.53)$$

$$f'(x_i) = \frac{-3f_i + 4f_{i+1} - f_{i+2}}{2h} + \frac{h^2}{3} f'''(\xi), \quad \xi \in (x_i, x_{i+2}), \quad (5.54)$$

$$f'(x_i) = \frac{f_{i+1} - f_{i-1}}{2h} - \frac{h^2}{6} f'''(\xi), \quad \xi \in (x_{i-1}, x_{i+1}), \quad (5.55)$$

$$f'(x_i) = \frac{f_{i-2} - 4f_{i-1} + 3f_i}{2h} + \frac{h^2}{3} f'''(\xi), \quad \xi \in (x_{i-2}, x_i), \quad (5.56)$$

$$f''(x_i) = \frac{f_{i+2} - 2f_{i+1} + f_i}{h^2} - h f'''(\xi), \quad \xi \in (x_i, x_{i+2}), \quad (5.57)$$

$$f''(x_i) = \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} - \frac{h^2}{12} f^{(4)}(\xi), \quad \xi \in (x_{i-1}, x_{i+1}), \quad (5.58)$$

$$f''(x_i) = \frac{f_i - 2f_{i-1} + f_{i-2}}{h^2} + h f'''(\xi) \quad \xi \in (x_{i-2}, x_i). \quad (5.59)$$

5.7 Podmíněnost numerické derivace a integrace

V oddílu 1.1 jsme se zabývali otázkou chyb v numerické matematice. Ty pochází mimo jiné z následujících tří zdrojů:

1. chyby numerického modelu, vzniklé obvykle diskretizací,
2. zaokrouhlovací chyby při výpočtu vstupních hodnot,
3. zaokrouhlovací chyby při aritmetických operacích.

Při numerickém výpočtu určitého integrálu nás především zajímaly chyby matematického modelu vzniklé diskretizací. Nepřesnosti funkčních hodnot nehrály takovou roli, protože jejich malé změny nemají významný vliv na hodnotu integrálu, jak jsme vysvětlili na začátku této kapitoly.

Při numerickém výpočtu derivací je však situace podstatně odlišná, zaokrouhlovací chyby mají podstatný vliv na celkovou chybu výpočtu. Všimneme si podrobněji této problematiky. Pro jednoduchost zanedbáme zaokrouhlovací chyby při aritmetických operacích. Budeme opět předpokládat, že uzly x_0, x_1, \dots, x_n jsou ekvidistantní s krokem $h > 0$ a označíme $\omega(x) = (x - x_0) \cdots (x - x_n)$.

Při numerickém derivování nahrazujeme funkci $f(x)$ interpolačním polynomem $L(x)$ s chybou $R(x)$, přičemž podle věty 4.4 pro dostatečně hladkou funkci $f(x)$ platí

$$f(x) = L(x) + R(x), \quad \text{kde} \quad R(x) = \omega(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

a $\xi \in \text{Int}\langle x_0, \dots, x_n, x \rangle$ je vhodné číslo. Polynom $L(x)$ pak derivujeme.

Výraz $R(x)$ představuje chybu numerického modelu při aproximaci funkce $f(x)$ pomocí $L(x)$. Samotné funkční hodnoty f_i figurují v $L(x)$, viz (4.3). Vyjádříme-li polynomy $L_i(x)$ jako polynomy v proměnné $t = \frac{x-x_0}{h}$, nezávisí jejich hodnoty na délce kroku h (srovnejte s (5.44)). Omezíme-li se na interpolaci, tj. $x \in \langle x_0, x_n \rangle$, zajímají nás hodnoty těchto polynomů pro $t \in \langle 0, n \rangle$. Není těžké přesvědčit se (např. pomocí vhodného programu CAS), že pro ne příliš velká n , což pro praktické použití stačí, nejsou hodnoty na tomto intervalu příliš velké. Proto při interpolaci funkčních hodnot $f(x)$ pomocí interpolačního polynomu $L(x)$ se malé chyby v f_i na celkové chybě nepodílí zásadně a rozhodující vliv má chyba $R(x)$.

To už není pravda, když rovnost $f(x) = L(x) + R(x)$ derivujeme. Ze vztahu (5.49) plyne, že k -tá derivace polynomu $L(x)$ je rovna k -té derivaci Newtonova interpolač-

ního polynomu vpřed $N^+(t)$ (což je jen jinak zapsaný Lagrangeův polynom) násobené zlomkem $1/h^k$. O derivacích polynomu $N^+(t)$ platí totéž, co bylo řečeno v předchozím odstavci — figurují v ní hodnoty f_i , které jsou násobené nějakými polynomy proměnné t , jejichž hodnoty nejsou na intervalu $\langle 0, n \rangle$ příliš velké. Malé chyby v hodnotách f_i by tedy nevadily. Ale vše je násobené faktorem $1/h^k$. Výsledek tudíž závisí na délce kroku h a pro malý krok se nepřesnosti ve funkčních hodnotách mohou zásadním způsobem zvětšit.

Naproti tomu chyba numerického modelu $R^{(k)}(x)$ se obecně zmenšuje s klesající délkou kroku h . Pro $k = 1$ a $k = 2$ je to vidět ze vzorců (5.34) a (5.35), ale platí to i pro vyšší derivace.

Celková chyba se tudíž skládá ze dvou částí — chyby numerického modelu, která je přímo úměrná nějaké mocnině délky kroku h , a chyby ve vstupních hodnotách, která je nepřímo úměrná nějaké mocnině h . Jejich chování je zcela protichůdné. Zatímco první část celkové chyby se zmenšujícím se h klesá, druhá neomezeně roste a naopak, první se zvětšujícím se h roste a druhá klesá. To je příčinou toho, že numerický výpočet derivací je špatně podmíněná úloha. Podrobnější rozbor celé problematiky lze nalézt například v [3, str. 92] nebo [43, str. 102].

Podíváme se podrobněji na nejjednodušší vzorec (5.52). Předpokládejme, že místo hodnot f_i a f_{i+1} použijeme hodnoty $\tilde{f}_i = f_i + \varepsilon_1$ a $\tilde{f}_{i+1} = f_{i+1} + \varepsilon_2$, kde $|\varepsilon_1| \leq \varepsilon$, $|\varepsilon_2| \leq \varepsilon$, přičemž $\varepsilon > 0$ je malé číslo charakterizující nepřesnost vstupních hodnot funkce $f(x)$. Za přibližnou hodnotu derivace $f'(x_i)$ tedy prohlásíme číslo

$$\tilde{f}'_i = \frac{\tilde{f}_{i+1} - \tilde{f}_i}{h} = \frac{f_{i+1} + \varepsilon_2 - f_i - \varepsilon_1}{h} = \frac{f_{i+1} - f_i}{h} + \frac{\varepsilon_2 - \varepsilon_1}{h}.$$

Určíme rozdíl mezi hodnotami \tilde{f}'_i a $f'(x_i)$. Podle (5.52) je

$$\tilde{f}'_i - f'(x_i) = \frac{\varepsilon_2 - \varepsilon_1}{h} + \frac{h}{2} f''(\xi), \quad \text{kde } \xi \in (x_i, x_{i+1}). \quad (5.60)$$

Protože obecně $\varepsilon_2 - \varepsilon_1 \neq 0$, je pro malé h část chyby způsobená nepřesnostmi vstupních hodnot velká, bude velká i relativní chyba a jde o špatně podmíněnou úlohu.

Odhadneme velikost celkové chyby. Předpokládejme, že $|f''(x)| \leq M$ na intervalu $\langle x_i, x_{i+1} \rangle$. Dostaneme

$$|\tilde{f}'_i - f'(x_i)| \leq \frac{|\varepsilon_2 - \varepsilon_1|}{h} + \frac{h}{2} |f''(\xi)| \leq \frac{2\varepsilon}{h} + \frac{Mh}{2} =: g(h).$$

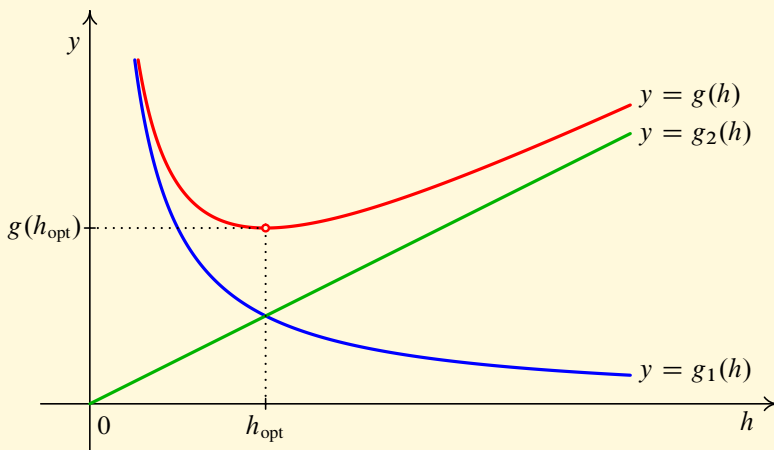
Vyšetříme průběh funkce $g(h)$ pro $h > 0$. Vypočteme její derivaci a určíme stacionární body. Je

$$g'(h) = -\frac{2\varepsilon}{h^2} + \frac{M}{2} = 0 \quad \Rightarrow \quad h_{\text{opt}} = 2\sqrt{\frac{\varepsilon}{M}}.$$

Jelikož $g''(h) = 4\varepsilon/h^3 > 0$, je funkce konvexní, má v h_{opt} globální minimum a klesá pro $0 < h \leq h_{\text{opt}}$ a roste pro $h \geq h_{\text{opt}}$. Hodnota globálního minima je $g(h_{\text{opt}}) = \sqrt{\varepsilon M}$.

Označme ještě $g_1(h) = 2\varepsilon/h$ a $g_2(h) = Mh/2$. Pak $g(h) = g_1(h) + g_2(h)$, přičemž první sčítanec je klesající funkce a druhý je rostoucí funkce. Grafy všech tří funkcí jsou znázorněny na obr. 5.14.

Existuje tudíž jakási optimální délka kroku h_{opt} . Volíme-li krok kratší, rychle narůstá část chyby způsobená nepřesností vstupních hodnot, volíme-li krok delší, narůstá chyba numerického modelu. Na základě podrobnějšího rozboru lze tvrdit, že mají-li hodnoty f_i a f_{i+1} určitý počet platných cifer, pak při volbě optimálního kroku je počet platných cifer aproximace f_i' přibližně poloviční, dochází tedy k podstatné ztrátě přesnosti (viz [13, str. 64]). Přitom lepší výsledek (jinou volbou délky kroku h) nemůžeme dosáhnout. Obdobná situace je i u jiných vzorců. Proto platí, že *numerická derivace je špatně podmíněnou úlohou*.

Obr. 5.14: Grafy funkcí $g(h)$, $g_1(h)$ a $g_2(h)$

Podstatně jiná je situace v případě numerické integrace pomocí kvadrurní formule tvaru (5.1). Ukážeme, že zaokrouhlovací chyby při výpočtu vstupních hodnot nemají zásadní vliv, pokud jsou všechny koeficienty A_i kladné. Opět pro jednoduchost nebudeme uvažovat zaokrouhlovací chyby při aritmetických operacích.

U jakékoliv prakticky použitelné kvadrurní formule očekáváme, že je přesná pro konstantní funkce. Volbou $f(x) = 1, x \in \langle a, b \rangle$, dostaneme ze vztahu (5.2), že

$$A_0 + A_1 + \cdots + A_n = \int_a^b dx = b - a.$$

Předpokládejme nyní, že místo přesných hodnot $f_i = f(x_i)$ použijeme hodnoty $\tilde{f}_i = f_i + \varepsilon_i, i = 0, 1, \dots, n$, přičemž $|\varepsilon_i| \leq \varepsilon$, kde $\varepsilon > 0$ je malé číslo charakterizující nepřesnost vstupních hodnot funkce $f(x)$. Za přibližnou hodnotu integrálu tedy prohlásíme číslo

$$\begin{aligned} Q(\tilde{f}) &= A_0 \tilde{f}_1 + A_1 \tilde{f}_2 + \cdots + A_n \tilde{f}_n = \\ &= A_0 f_0 + A_1 f_1 + \cdots + A_n f_n + A_0 \varepsilon_0 + A_1 \varepsilon_1 + \cdots + A_n \varepsilon_n = \\ &= Q(f) + A_0 \varepsilon_0 + A_1 \varepsilon_1 + \cdots + A_n \varepsilon_n. \end{aligned}$$

Odhadneme rozdíl mezi hodnotami $Q(\tilde{f})$ a $Q(f)$. Vyjde

$$\begin{aligned} |Q(\tilde{f}) - Q(f)| &= |A_0\varepsilon_0 + A_1\varepsilon_1 + \cdots + A_n\varepsilon_n| \leq \\ &\leq |A_0||\varepsilon_0| + |A_1||\varepsilon_1| + \cdots + |A_n||\varepsilon_n| \leq \\ &\leq (A_0 + A_1 + \cdots + A_n)\varepsilon = \varepsilon(b - a). \end{aligned}$$

Využili jsme přitom předpokladu, že koeficienty A_i jsou kladné. Předchozí odhad nezávisí ani na počtu uzlových bodů, ani na délce kroků. Pro malé ε je tudíž vliv nepřesností vstupních hodnot nepodstatný, rozhodující podíl na celkové chybě má chyba numerického modelu. To je ve shodě s tím, co jsme o výpočtu určitého integrálu zmínili na začátku této kapitoly.

Předchozí zdůvodnění ukazuje, proč je pro dobrou podmíněnost numerické integrace důležité, aby koeficienty kvadrurní formule byly kladné. Tento požadavek splňují např. Gaussovy-Legendrovy formule, viz vzorec (5.24). Jiná je situace u Newtonových-Cotesových formulí. Lze ukázat, že pro uzavřené formule jsou všechny koeficienty kladné pouze pro $n = 1, 2, \dots, 7$ a $n = 9$, viz [43, str. 179]. Je to další důvod, proč není vhodné používat tyto formule pro vysoká n .

Pojmy k zapamatování

- kvadrurní formule
- chyba kvadrurní formule
- stupeň přesnosti kvadrurní formule
- uzavřené Newtonovy-Cotesovy kvadrurní formule
- lichoběžníkové, Simpsonovo, Simpsonovo 3/8 a Boolovo pravidlo
- otevřené Newtonovy-Cotesovy kvadrurní formule
- obdélníkové, otevřené lichoběžníkové a Milnovo pravidlo
- složené Newtonovy-Cotesovy kvadrurní formule
- Gaussovy kvadrurní formule
- Legendrovy polynomy
- Gaussovy-Legendrovy kvadrurní formule
- Rombergova kvadratura
- numerická derivace
- vliv zaokrouhlovacích chyb na numerickou derivaci a integraci

Kontrolní otázky

1. Proč je nutné (resp. vhodné) v některých případech použití numerické integrace?
2. Jaký je princip numerického integrování?
3. Napište obecný tvar kvadraturního formule.
4. Jak je definován stupeň přesnosti kvadraturní formule?
5. Jaký maximální stupeň přesnosti může mít kvadraturní formule používající $n + 1$ uzlů?
6. Napište, jakým typem funkce nahrazujeme integrand při použití uzavřených Newtonových-Cotesových formulí pro $n = 1, 2$.
7. Napište, jakým typem funkce nahrazujeme integrand při použití otevřených Newtonových-Cotesových formulí pro $n = 2, 3, 4$.
8. V čem spočívá rozdíl při použití uzavřených a otevřených Newtonových-Cotesových formulí?
9. Pro jaký počet uzlů n se nejčastěji používají Newtonovy-Cotesovy formule?
10. Proč se nepoužívají Newtonovy-Cotesovy formule pro velká n ?
11. Popište, případně načrtněte princip lichoběžníkového pravidla.
12. Popište, případně načrtněte princip Simpsonova pravidla.
13. Pro polynomy jakého stupně je přesné lichoběžníkové pravidlo?

14. Pro polynomy jakého stupně je přesné Simpsonovo pravidlo?
15. Proč používáme při integraci častěji složené kvadraturní formule, a ne základní kvadraturní formule?
16. Vysvětlete princip složených kvadraturních formulí.
17. Načrtněte princip složeného Simpsonova pravidla.
18. Jak musíme volit uzly kvadraturní formule, abychom dosáhli maximálního možného stupně přesnosti?
19. Co jsou to Gaussovy-Legendrovy kvadraturní formule?
20. Jak najdeme koeficienty Gaussových-Legendrových kvadraturních formulí?
21. Jaký je rozdíl mezi Newtonovými-Cotesovými a Gaussovými kvadraturními formullemi?
22. Jakou výhodu má použití Gaussových kvadraturních formulí oproti Newtonovým-Cotesovým formulím?
23. Na jakém principu je založena Rombergova metoda?
24. Která z probíraných metod pro numerickou integraci nepožaduje použití ekvidistantních uzlů?
25. Na jakém principu je založené numerické derivování?
26. Jak lze vyjádřit chybu modelu při numerickém derivování?
27. Vysvětlete, proč je numerická derivace špatně podmíněná úloha. Jak je tomu u integrace?

Testy ke kapitole 5

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Obecná kvadraturní formule se čtyřmi ekvidistantními uzly využívající oba krajní body integračního oboru, která aproximuje integrál $\int_{-2}^1 f(x) dx$, má tvar

$$Q(f) = f(-2) + f(-1) + f(0) + f(1).$$

$$Q(f) = A_0 f(-2) + A_1 f(-1) + A_2 f(0) + A_3 f(1).$$

$$Q(f) = A_0 f(0) + A_1 f(1) + A_2 f(2) + A_3 f(3).$$

$$Q(f) = A_0(f_{-2}) + A_1(f_{-1}) + A_2(f_0) + A_3(f_1).$$

2. (1b.) Kvadrurní formule $Q(f) = \sum_{i=0}^n A_i f(x_i)$ má stupeň přesnosti N , jestliže

$R(1) = \dots = R(x^N) = 0$ a $R(x^{N+1}) \neq 0$, kde $R(f)$ je vzorec pro chybu kvadrurní formule.

$R(1) = \dots = R(x^{N-1}) = 0$ a $R(x^N) \neq 0$, kde $R(f)$ je vzorec pro chybu kvadrurní formule.

používá $N + 1$ uzlových bodů.

má N koeficientů $A_i, i = 0, \dots, N - 1$.

3. (1b.) Mezi otevřené Newtonovy-Cotesovy kvadrurní formule nepatří
 obdélíkové pravidlo, otevřené lichoběžníkové pravidlo,
 Simpsonovo pravidlo, Milnovo pravidlo.

4. (1b.) Uzavřená Newtonova-Cotesova kvadrurní formule řádu $n = 3$ používající co nejmenší počet uzlů má
- 4 uzlové body a nahrazuje funkci $f(x)$ polynomem 3. stupně,
 - 3 uzlové body a nahrazuje funkci $f(x)$ parabolou,
 - 2 uzlové body a nahrazuje funkci $f(x)$ parabolou,
 - 2 uzlové body a nahrazuje funkci $f(x)$ přímkou.
5. (1b.) V případě, že zvolíme pevně uzly $x_i, i = 0, \dots, n$, nějaké kvadrurní formule, která má stupeň přesnosti alespoň n , pak
- koeficienty A_i jsou určeny jednoznačně, nezávisí na nahrazované funkci $f(x)$.
 - koeficienty závisí na nahrazované funkci $f(x)$.
 - koeficienty A_i volíme tak, aby jejich součet byl roven jedné.
 - koeficienty volíme tak, aby $A_i = A_{n-i}, i = 0, \dots, n$.

6. (1b.) Je možné pomocí obdélníkového pravidla nahradit polynom 1. stupně tak, aby byla chyba nulová?

Ano, ale jen když je grafem polynomu vodorovná přímka.

Ne. Polynom 1. stupně má za graf přímku, která obecně nemůže splývat s grafem konstantní funkce, kterou danou funkci nahrazujeme při použití obdélníkového pravidla.

Ano. Pro lichý počet $n + 1$ uzlových bodů souměrných kolem středu integračního intervalu je stupeň přesnosti alespoň $n + 1$.

Ne. Stupeň přesnosti je vždy o jedničku menší než počet použitých uzlových bodů.

7. (1b.) Které kvadraturní formule dosahují při daném počtu uzlových bodů maximálního stupně přesnosti?

Uzavřené Newtonovy-Cotesovy
formule.

Rombergovy formule.

Otevřené Newtonovy-Cotesovy
formule.

Gaussovy-Legendrovy formule.

8. (1b.) Při konstrukci složených kvadraturních formulí obvykle nepoužíváme Gaussovy formule, protože

je obtížné odhadnou velikost jejich chyby.

takto vytvořené metody nejsou numericky stabilní.

nepoužívají ekvidistantní uzly, což je nevýhodné, když chceme např. zdvojnásobit počet intervalů.

mají nízký stupeň přesnosti.

9. (1b.) Kvadraturní formule získaná integrací interpolačního polynomu a používající lichý počet $n + 1$ uzlových bodů, které jsou rozloženy souměrně kolem středu intervalu $\langle a, b \rangle$, má stupeň přesnosti

alespoň $n + 1$.

alespoň $n + 2$.

alespoň $2n + 1$.

alespoň $2n + 2$.

10. (1b.) O kořenech Legendrových polynomů platí:

Všechny jsou reálné, jednoduché, jsou symetrické vzhledem k počátku a alespoň dva leží vně otevřeného intervalu $(-1, 1)$.

Všechny jsou reálné, jednoduché, jsou symetrické vzhledem k počátku a leží v otevřeném intervalu $(-1, 1)$.

Aspoň jeden je komplexní s nenulovou imaginární částí a všechny leží uvnitř jednotkového kruhu se středem v počátku.

Všechny jsou reálné, jsou symetrické vzhledem k počátku a počet vícenásobných kořenů je lichý.

11. (1b.) Na čem je založeno numerické derivování funkce $f(x)$?

Na náhradě funkce $f(x)$ kubickým splajnem.

Na náhradě funkce $f(x)$ polynomem získaným metodou nejmenších čtverců.

Na náhradě funkce $f(x)$ interpolačním polynomem.

Na sestavení vhodné diferenciální rovnice pro funkci $f(x)$.

12. (1b.) Numerická derivace je

obecně špatně podmíněná úloha.

obecně velmi dobře podmíněná úloha.

je dobře podmíněná úloha, když počítáme první derivaci.

13. (1b.) Která vlastnost zaručuje, že je kvadrurní formule $Q(f) = A_0 f(x_0) + \dots + A_n f(x_n)$ dobře podmíněná?

Všechny koeficienty A_i jsou celočíselné.

Všechny koeficienty A_i jsou stejné.

Všechny koeficienty A_i jsou menší než jedna.

Všechny koeficienty A_i jsou kladné.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

3. (1b.) Newtonovy-Cotesovy kvadrurní formule, které zahrnují mezi uzly i krajní body intervalu $\langle a, b \rangle$, tj. $x_0 = a$ a $x_n = b$, se nazývají
otevřené. úplné. uzavřené. periodické.
4. (1b.) V případě, že jsme použili lichoběžníkové pravidlo, pak
jsme funkci $f(x)$ na intervalu $\langle a, b \rangle$ nahradili přímkou rovnoběžnou s osou x ,
jsme interval $\langle a, b \rangle$ rozdělili na dva podintervaly a na každém jsme funkci $f(x)$
nahradili přímkou, přičemž tyto přímky nesplyvají.
jsme funkci $f(x)$ na intervalu $\langle a, b \rangle$ nahradili přímkou,
5. (1b.) Vyberte nepravdivou odpověď: Mezi kvadrurní formule, jejichž koeficienty A_i získáme integrací interpolačního polynomu, který je určen body $[x_i, f(x_i)]$, $i = 0, \dots, n$, patří
Newtonovy-Cotesovy kvadrurní formule, Gaussovy kvadrurní formule,
Lagrangeovy kvadrurní formule.

6. (1b.) Je možné pomocí Simpsonova pravidla nahradit polynom 3. stupně tak, aby byla chyba nulová?

Ano, ale jen když je grafem polynomu parabola.

Ne. Polynom 3. stupně má za graf kubickou křivku, která obecně nemůže splývat s grafem kvadratické funkce, tj. parabolou, kterou danou funkci nahrazujeme při použití Simpsonova pravidla.

Ne. Stupeň přesnosti je vždy o jedničku menší než počet použitých uzlových bodů.

Ano. Pro lichý počet $n + 1$ uzlových bodů souměrných kolem středu integračního intervalu je stupeň přesnosti alespoň $n + 1$.

7. (1b.) Rombergova kvadratura je založena na

asymptotickém vzorci pro složené otevřené lichoběžníkové pravidlo.

asymptotickém vzorci pro složené lichoběžníkové pravidlo.

asymptotickém vzorci pro složené obdélníkové pravidlo.

asymptotickém vzorci pro složené Simpsonovo pravidla.

8. (1b.) Při konstrukci složených kvadraturních formulí obvykle nepoužíváme otevřené Newtonovy-Cotesovy formule, protože

mají nízký stupeň přesnosti.

je obtížné odhadnou velikost jejich chyby.

takto vytvořené metody nejsou numericky stabilní.

nepoužívají krajní body intervalu, což je nevýhodné když chceme např. zdvojnásobit počet intervalů.

9. (1b.) Kvadraturní formule používající $n + 1$ uzlových bodů, může mít stupeň přesnosti
nejvýše $n - 1$. nejvýše $2n - 1$. nejvýše $2n$.

nejvýše $2n + 1$. nejvýše n . nejvýše $n + 1$.

10. (1b.) Má-li kvadrurní formule používající $n + 1$ uzlů stupeň přesnosti aspoň n , musí se uzly shodovat s kořeny některého Legendrova polynomu.

nejsou její koeficienty určeny jednoznačně.

lze její koeficienty získat integrací speciálních Lagrangeových polynomů $L_i(x)$ určených uzly formule.

musí se jednat o některou Newtonovu-Cotesovu kvadrurní formuli.

11. (1b.) Proč je numerická derivace špatně podmíněná úloha?

Protože s rostoucí délkou kroku h roste chyba numerického modelu.

Protože s klesající délkou kroku h roste chyba numerického modelu.

Protože s klesající délkou kroku h klesá chyba způsobená nepřesnostmi vstupních hodnot.

Protože s klesající délkou kroku h roste chyba způsobená nepřesnostmi vstupních hodnot.

12. (1b.) Nechť funkce $f(x)$ a $g(x)$ mají na intervalu $\langle a, b \rangle$ spojitou první derivaci. Vyberte pravdivé tvrzení.

Jestliže se málo liší jejich určité integrály $\int_a^b f(x) dx$ a $\int_a^b g(x) dx$, liší se málo i jejich derivace.

Jestliže se $f(x)$ a $g(x)$ málo liší, mohou se jejich derivace lišit hodně.

Jestliže se $f(x)$ a $g(x)$ málo liší, mohou se jejich určité integrály $\int_a^b f(x) dx$ a $\int_a^b g(x) dx$ lišit hodně.

Jestliže se $f(x)$ a $g(x)$ málo liší, liší se málo i jejich derivace.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Kapitola 6

Numerické řešení obyčejných diferenciálních rovnic

Cíle

Po prostudování této kapitoly budete schopni vysvětlit:

- co je to obyčejná diferenciální rovnice (ODR) prvního řádu v implicitním resp. explicitním tvaru a co je její řešení,
- jak vypadá Cauchyova počáteční úloha pro ODR prvního řádu,
- proč je nutné hledat řešení počáteční úlohy numericky a jaký je princip numerického řešení,

- jaký je rozdíl mezi jednokrokovými a vícekrokovými metodami,
- jaký je obecný tvar jednokrokové explicitní a implicitní metody,
- co je to lokální a globální chyba, řád, konvergence a A-stabilita jednokrokové metody,
- co jsou to metody Rungeho-Kutty a jaký je jejich princip,
- jaký je obecný tvar lineární vícekrokové explicitní a implicitní metody,
- co je to lokální a globální chyba, řád, konvergence a A-stabilita lineární vícekrokové metody,
- co jsou Adamsovy-Bashforthovy a Adamsovy-Moultonovy metody a metody zpětného derivování a jaký je princip jejich odvození,
- co jsou metody prediktor-korektor,
- čemu se říká tuhé problémy a na co je nutné dbát při jejich řešení,
- jak se výsledky pro jednu ODR prvního řádu přenesou na soustavy ODR prvního řádu,
- jak přepíšeme jednu ODR vyššího řádu na soustavu ODR prvního řádu.

Jde o jednu z nejdůležitějších úloh numerické matematiky, protože diferenciální rovnice patří k nejvýznamnějším matematickým modelům reálných problémů. Přitom pouze malá část obyčejných diferenciálních rovnic má řešení, které lze vyjádřit v uzavřeném tvaru, tj. pomocí elementárních funkcí.

6.1 Obyčejné diferenciální rovnice 1. řádu

6.1.1 Základní vlastnosti obyčejných diferenciálních rovnic 1. řádu

Připomeneme si základní poznatky z tzv. *kvalitativní teorie* obyčejných diferenciálních rovnic. Podrobněji viz [34].

- Rovnice $F(x, y, y') = 0$ se nazývá *obyčejná diferenciální rovnice prvního řádu v implicitním tvaru* s neznámou funkcí $y(x)$ nezávisle proměnné x . Příkladem takové rovnice je

$$\ln(x^2 - yy' + y^2) - (y')^3 + x \sin(xy y') - \sqrt{y - y'} = 0.$$

- Pro vyšetřování rovnic je důležitý případ, kdy se dá osamostatnit první derivace y' , tj. rovnice má tvar

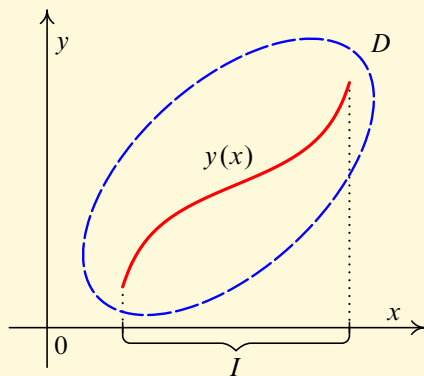
$$y' = f(x, y). \quad (6.1)$$

O takové rovnici říkáme, že je v *explicitním tvaru*. O funkci $f(x, y)$ předpokládáme, že je definována na nějaké otevřené množině $D \subset \mathbb{R}^2$.

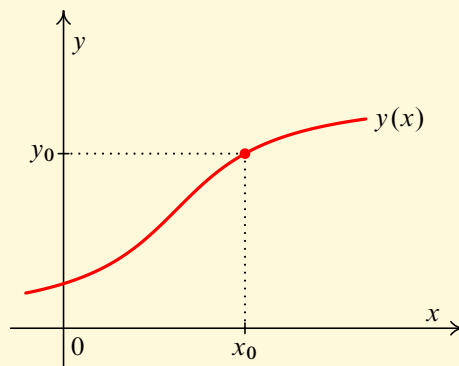
Příklady takových rovnic jsou

$$y' = x^2 + y^2, \quad y' = \frac{y}{x}, \quad y' = y^2, \quad y' = \sin xy, \quad y' = \sqrt{x^2 - y^2} \quad \text{apod.}$$

- *Řešením* je funkce $y(x)$ definovaná na *intervalu* I , která splňuje rovnici (6.1). Musí tedy platit, že $[x, y(x)] \in D$ pro $x \in I$ (jinými slovy, graf funkce $y(x)$ leží v množině D — viz obr. 6.1 a) a $y'(x) = f(x, y(x))$ pro každé $x \in I$.
- V každém bodě $[x, y(x)]$ grafu řešení tedy platí, že tečna ke grafu v tomto bodě má směrnici rovnou číslu $f(x, y(x))$. Protože funkci $f(x, y)$ známe, můžeme sestrojít (teoreticky) v každém bodě $[x, y]$ vázaný vektor, jehož směrnice je $f(x, y)$. Graf řešení se musí v každém bodě, kterým prochází, dotýkat tohoto vektoru. Tyto vektory tvoří *směrové pole* diferenciální rovnice. Na obr. 6.2, 6.3 a 6.4 jsou znázorněna směrová pole



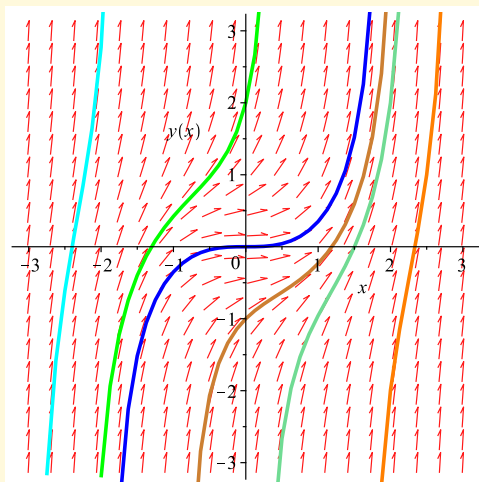
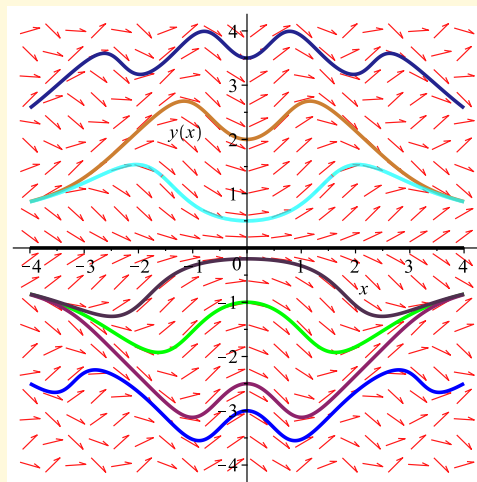
a) Řešení diferenciální rovnice



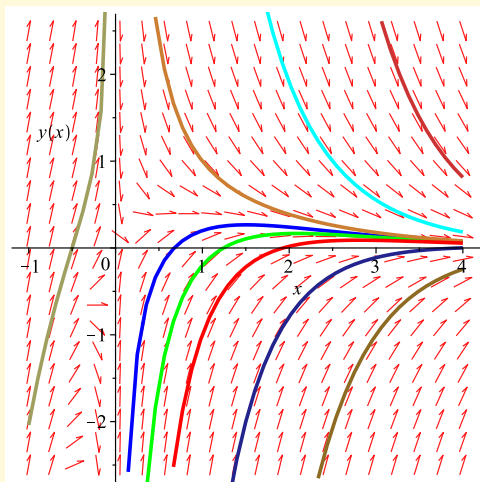
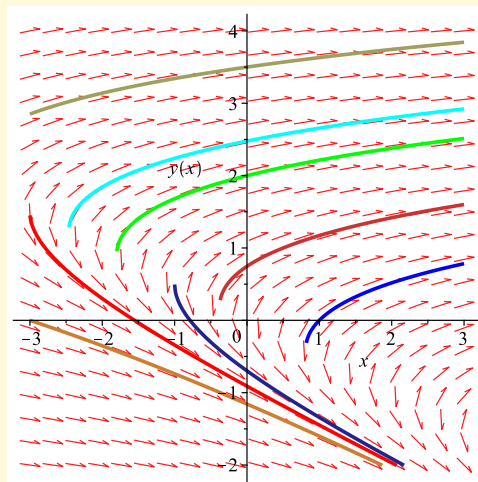
b) Cauchyova počáteční úloha

Obr. 6.1

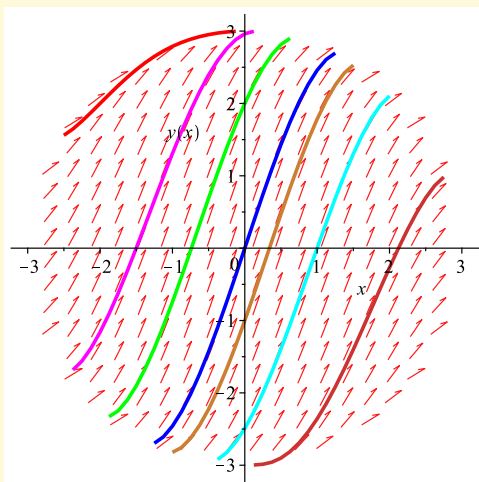
a několik řešení šesti diferenciálních rovnic. Všimněte si, že u rovnic na obr. 6.3 a) až 6.4 b) není množina D rovna celé rovině \mathbb{R}^2 (určete definiční obory pravých stran těchto rovnic).

a) $y' = x^2 + y^2$ b) $y' = \sin(xy)$

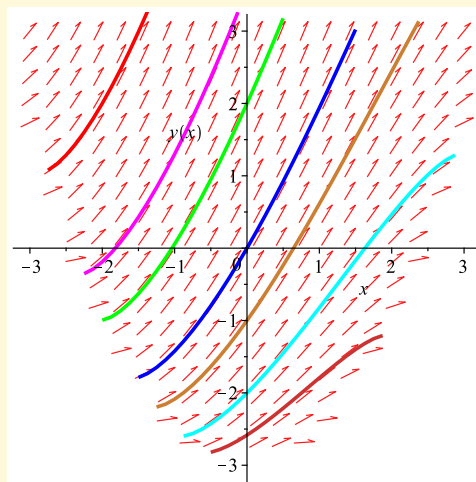
Obr. 6.2: Směrová pole a řešení diferenciálních rovnic — část 1

a) $y' = 2e^{-x} - (1 + 1/x)y$ b) $y' = 1/(x + 2y)$

Obr. 6.3: Směrová pole a řešení diferenciálních rovnic — část 2



$$\text{a) } y' = \sqrt{9 - x^2 - y^2}$$



$$\text{b) } y' = \sqrt{y - x^2/2 + 3}$$

Obr. 6.4: Směrová pole a řešení diferenciálních rovnic — část 3

- Rovnice $y' = f(x, y)$ má obvykle nekonečně mnoho řešení. Hledáme tedy řešení, které splňuje nějakou dodatečnou podmínku. Tato podmínka by měla jednoznačně určit jediné řešení.

My budeme hledat řešení, jehož graf prochází zadaným bodem $[x_0, y_0]$. Chceme tedy, aby platila tzv. *počáteční podmínka* $y(x_0) = y_0$ — viz obr. 6.1 b). Úloha, kterou budeme řešit, má tudíž tvar

$$\begin{aligned} y'(x) &= f(x, y(x)), \\ y(x_0) &= y_0. \end{aligned} \tag{6.2}$$

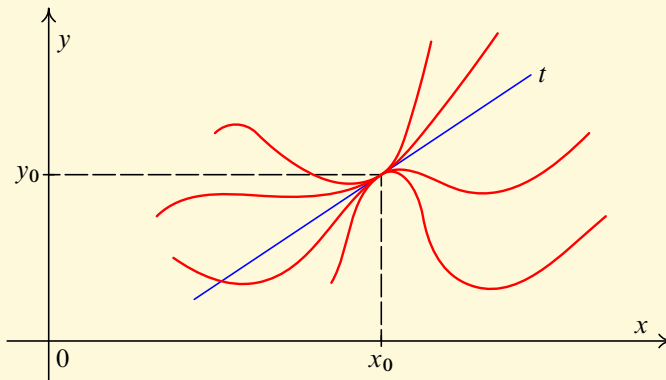
Říká se jí obvykle *Cauchyova počáteční úloha*.

- Připomeňme si, jaké vlastnosti funkce $f(x, y)$ zaručují, že Cauchyova počáteční úloha (6.2) má řešení a že toto řešení je jediné.

1) Z teorie obyčejných diferenciálních rovnic je známo, že *spojitost funkce* $f(x, y)$ zaručuje existenci řešení počáteční úlohy.

Toto řešení ale nemusí být jediné. Pak říkáme, že v bodě $[x_0, y_0]$ je porušena jednoznačnost. Protože všechna řešení počáteční úlohy (6.2) musí mít v bodě $[x_0, y_0]$

společnou tečnu t , musí dojít k jakémusi „rozvětvení“ — viz obr. 6.5. Numerické hledání řešení je pak velmi problematické.



Obr. 6.5: Porušení jednoznačnosti v bodě $[x_0, y_0]$

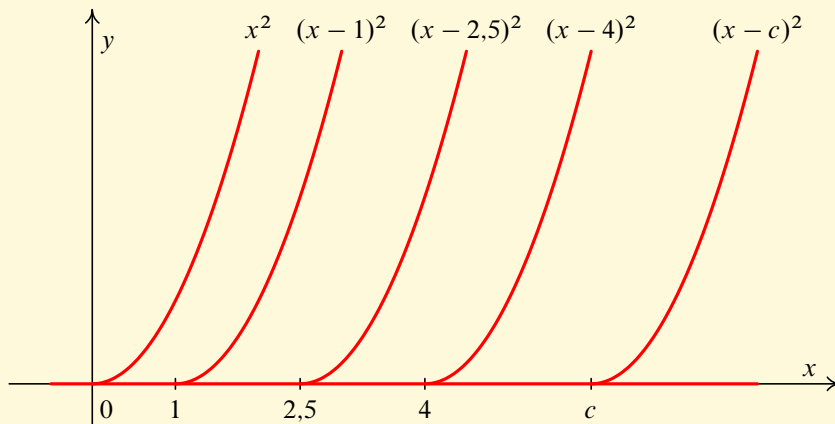
Např. počáteční úloha $y' = 2\sqrt{|y|}$, $y(0) = 0$, má spojitou pravou stranu $f(x, y) = 2\sqrt{|y|}$ v množině $D = \mathbb{R}^2$. Uvedená úloha má ale nekonečně mnoho řešení.

Jedním je nulová funkce $y(x) \equiv 0$, dalšími řešeními jsou funkce

$$y_c = \begin{cases} (x - c)^2, & x \geq c, \\ 0, & x < c, \end{cases}$$

kde $c \in \mathbb{R}_0^+$ je libovolné číslo. Řešení jsou znázorněna na obr. 6.6. Že jde opravdu o řešení, se snadno ověří dosazením do rovnice.

- 2) Standardní vlastností, která zaručuje, že počáteční úloha (6.2) má jediné řešení, je *Lipschitzova podmínka* (existuje konstanta $L > 0$ taková, že je v množině D splněna nerovnost $|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|$; stačí, aby tato vlastnost platila lokálně). K platnosti této podmínky stačí, aby v okolí bodu $[x_0, y_0]$ byla *spojitá parciální derivace* $\frac{\partial}{\partial y} f(x, y)$. Tuto podmínku rovnice, se kterými se v aplikacích setkáváme, většinou splňují.

Obr. 6.6: Řešení úlohy $y' = 2\sqrt{|y|}$, $y(0) = 0$

6.1.2 Numerické řešení diferenciálních rovnic 1. řádu

Budeme hledat řešení $y(x)$ rovnice $y' = f(x, y)$ splňující počáteční podmínku $y(x_0) = y_0$ na intervalu $\langle a, b \rangle$, kde $a = x_0$ — viz obr. 6.7 a). Jde tedy o řešení počáteční úlohy (6.2) vpravo od bodu x_0 . (Podobně bychom mohli hledat řešení vlevo od tohoto bodu.) Zajímá nás hodnota $y(b)$. Musíme mít ovšem zaručeno, že řešení na intervalu $\langle a, b \rangle$ opravdu existuje!

Příklad 6.1 Najděte řešení počáteční úlohy $y' = y^2$, $y(2) = 1$.

Řešení. Jde o obyčejnou diferenciální rovnici se separovanými proměnnými. Zřejmě platí $f(x, y) = y^2$, $\frac{\partial f}{\partial y} = 2y$. Za množinu D lze volit celou rovinu \mathbb{R}^2 . Protože funkce f i její parciální derivace jsou na D spojité, prochází každým bodem roviny právě jedno řešení.

Rovnice má zřejmě řešení $y \equiv 0$. Pro $y \neq 0$ separací a integrací vyjde:

$$\frac{dy}{dx} = y^2, \quad \int \frac{dy}{y^2} = \int dx, \quad y = -\frac{1}{x + c}.$$

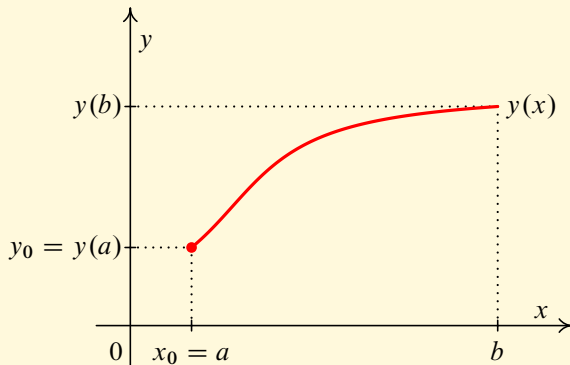
$$\frac{dy}{y^2} = dx, \quad -\frac{1}{y} = x + c, \quad c \in \mathbb{R},$$

Z počáteční podmínky $y(2) = 1$ dostaneme:

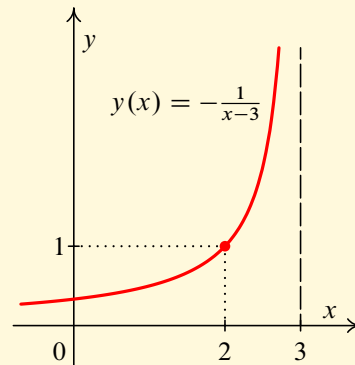
$$1 = -\frac{1}{2+c} \quad \Rightarrow \quad c = -3 \quad \Rightarrow \quad y(x) = -\frac{1}{x-3}$$

Řešení (větev rovnoosé hyperboly) je znázorněno na obr. 6.7 b). Jeho maximálním definičním oborem je interval $(-\infty, 3)$. Tedy nelze hledat řešení např. na intervalu $\langle 2; 4 \rangle$! ▲

V podobných případech počítačový program občas najde nějaké „řešení“, ale pochopitelně se jedná o nesmysl. Je na uživateli, aby takovou věc odhalil, výstupu z počítače nelze slepě věřit.



a) Cauchyova počáteční úloha —
řešení na intervalu $\langle a, b \rangle$



b) Řešení počáteční úlohy
 $y' = y^2, y(2) = 1$

Obr. 6.7

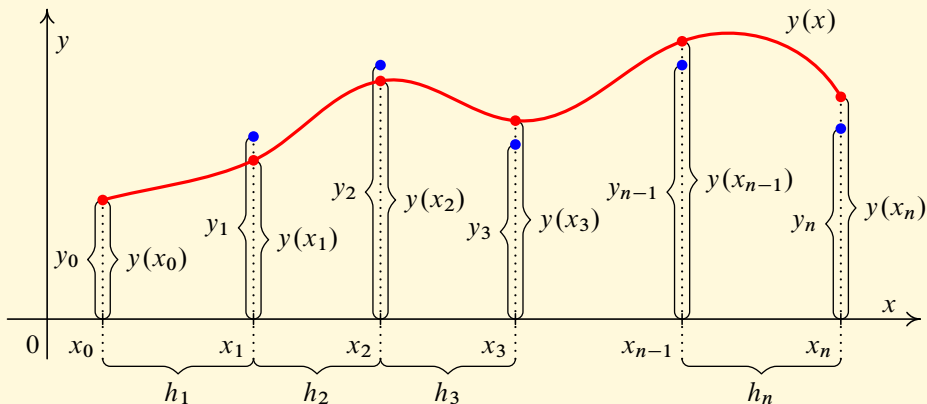
Princip numerického řešení

Zvolme body $a = x_0 < x_1 < \dots < x_n = b$. Numerickým řešením počáteční úlohy (6.2) rozumíme nalezení přibližných hodnot $y(x_i)$, které označíme y_i . Body x_i nazýváme *uzlové body* nebo *uzly*. Zdůrazněme, že obecně platí pouze $y(x_i) \approx y_i$. Jde tudíž o nalezení tabulky hodnot

$$\begin{array}{c|c|c|c|c} x_i & x_0 & x_1 & \dots & x_n \\ \hline y_i & y_0 & y_1 & \dots & y_n \end{array}.$$

Výsledek je znázorněn na obr. 6.8.

Označme $x_i - x_{i-1} = h_i$, $i = 1, \dots, n$. Čísla h_i nazýváme *kroky*. Obecně nemusí být stejné. Pokud jsou stejné, nazývá se dělení *ekvidistantní* a platí $h_i = h = \frac{b-a}{n}$.



Obr. 6.8: Princip numerického řešení počáteční úlohy $y' = f(x, y)$, $y(x_0) = y_0$; $y(x)$ je přesné řešení, jehož analytický vzorec obecně neznáme.

Poznámka 6.2

- i) Pochopitelně bychom nejradyji měli analytický vzorec řešení. Řešení ale ve většině případů nelze popsat elementárními funkcemi. Proto se musíme uskrovnit. Místo vzorce dostaneme numerickým řešením hodnoty řešení $y(x)$ pouze na konečné množině uzlových bodů a k tomu ještě jen přibližně. Nic lepšího ale najít neumíme.
- ii) I když nás v některých případech zajímá pouze hodnota $y(b)$, musíme najít hodnoty i v ostatních uzlových bodech. Najít přímo $y(b)$ neumíme. Naopak lze čekat, že čím více uzlových bodů zvolíme mezi a a b a čím menší budou kroky, tím lepší přibližnou hodnotu pro $y(b)$ dostaneme. O tom budeme mluvit v dalším textu.

Klasifikace metod

Metody dělíme na *jednokrokové* a *vícekrokové*. Připomeňme, že hodnota $y(x_0)$ je známá a je rovna číslu y_0 ; to říká počáteční podmínka $y(x_0) = y_0$.

Jednokrokové metody

Hodnota y_{i+1} se počítá z jediné hodnoty y_i , $i = 0, \dots, n - 1$, která odpovídá bezprostředně předcházejícímu uzlovému bodu. Tedy

- ze známé hodnoty y_0 počítáme y_1 ,
- z y_1 pak počítáme y_2 ,
- z y_2 pak počítáme y_3 atd.

Vícekové metody

K výpočtu hodnoty y_{i+1} , $i = 0, \dots, n-1$, se používá více hodnot, které odpovídají předcházejícím uzlovým bodům. Obecná k -kroková metoda, kde $k \in \mathbb{N}$, $k \geq 2$, počítá y_{i+1} z k hodnot $y_i, y_{i-1}, \dots, y_{i-k+1}$, $i = k-1, \dots, n-1$, odpovídajících předcházejícím k uzlovým bodům. Např. pro $k = 3$, tj. tříkrokovou metodou,

- ze tří hodnot y_0, y_1 a y_2 počítáme y_3 ,
- ze tří hodnot y_1, y_2 a y_3 počítáme y_4 ,
- ze tří hodnot y_2, y_3 a y_4 počítáme y_5 atd.

Tedy obecně y_{i+1} počítáme ze tří hodnot y_{i-2}, y_{i-1} a y_i , $i = 2, \dots, n-1$.

Zadání počáteční úlohy však obsahuje jen hodnotu y_0 . Abychom mohli odstartovat některou k -krokovou hodnotu, kde $k \geq 2$, musíme nějak určit hodnoty y_1, y_2, \dots, y_{k-1} . K tomu se použije nějaká jednokroková metoda. Tyto hodnoty *musíme určit co nejpřesněji!* „Pokážeme-li“ numerické řešení hned na začátku, je zbytečné potom nasazovat nějakou sofistikovanou vícekrokovou metodu a počítat další hodnoty z výchozích špatných hodnot.

Konkrétními vzorci pro jednokrokové i vícekrokové metody se budeme zabývat v dalších odstavcích.

6.1.3 Jednokrokové metody

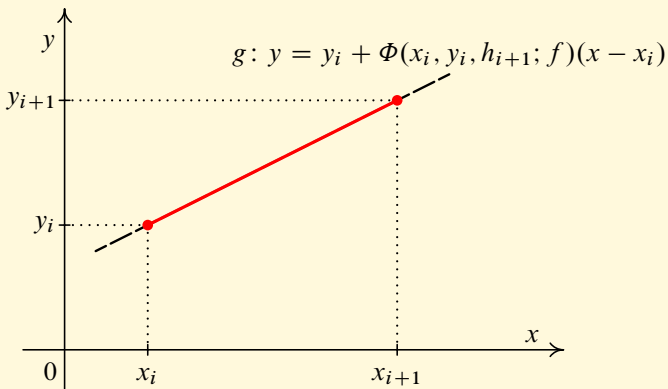
Připomeňme, že řešíme počáteční úlohu $y' = f(x, y)$, $y(x_0) = y_0$, na intervalu $\langle a, b \rangle$. Přitom $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$, $h_i = x_i - x_{i-1}$, kde $i = 1, \dots, n$, jsou kroky a y_i jsou přibližné hodnoty řešení $y(x)$ v uzlových bodech x_i , tj. $y_i \approx y(x_i)$, $i = 1, \dots, n$. Metody rozdělíme na dva typy: *explicitní* a *implicitní*.

Obecná explicitní jednokroková metoda se zapisuje ve tvaru

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f), \quad i = 0, \dots, n-1, \quad (6.3)$$

kde Φ je funkce tří proměnných závisující též na funkci f .

Geometricky lze vzorec (6.3) interpretovat takto: Bodem o souřadnicích $[x_i, y_i]$ proložíme přímkou o směrnici $\Phi(x_i, y_i, h_{i+1}; f)$. Její rovnice ve směrnicevém tvaru bude $y = g(x)$, kde $g(x) = y_i + \Phi(x_i, y_i, h_{i+1}; f)(x - x_i)$. Hodnota funkce $g(x)$ v bodě $x_{i+1} = x_i + h_{i+1}$ bude $g(x_{i+1}) = y_i + \Phi(x_i, y_i, h_{i+1}; f)(x_{i+1} - x_i) = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f) = y_{i+1}$ — viz obr. 6.9. Podle toho, jak volíme funkci $\Phi(x_i, y_i, h_{i+1}; f)$, dostáváme různé metody.



Obr. 6.9: Obecná explicitní jednokroková metoda

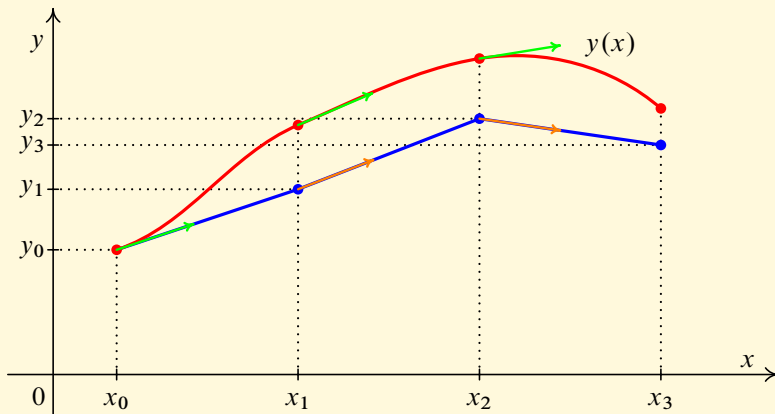
Explicitní Eulerova metoda

Nejjednodušší explicitní jednokrokovou metodou je *explicitní Eulerova¹ metoda*

$$y_{i+1} = y_i + h_{i+1}f(x_i, y_i), \quad i = 0, \dots, n-1. \quad (6.4)$$

Obvykle se volí konstantní krok $h = \frac{b-a}{n}$, tedy uzlové body jsou ekvidistantní. V tomto případě je $\Phi(x_i, y_i, h_{i+1}; f) = f(x_i, y_i)$, takže za směrnici pomocné přímky se volí hodnota daná směrovým polem rovnice $y' = f(x, y)$ v bodě $[x_i, y_i]$ — viz obr. 6.10. Lomená čára spojující body $[x_0, y_0], [x_1, y_1], \dots, [x_n, y_n]$, kterou takto dostaneme, se nazývá *Eulerův polygon*. Tato metoda je velmi jednoduchá, ale nepřesná.

¹**Leonhard Euler** (1707–1783) (čti ojler) — švýcarský matematik, fyzik, mechanik a astronom, jeden z největších matematiků všech dob. Napsal kolem 850 prací, ovlivnil všechny tehdejší disciplíny. Od roku 1766 byl slepý, svá díla diktoval. Zavedl např. označení pro konstanty π , e a i .



Obr. 6.10: Explicitní Eulerova metoda pro $n = 3$;
zelené šipky určují směrové pole v bodech $[x_i, y(x_i)]$,
hnědé šipky určují směrové pole v bodech $[x_i, y_i]$

Obecná implicitní jednokroková metoda se zapisuje ve tvaru

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}, y_{i+1}; f), \quad i = 0, \dots, n-1, \quad (6.5)$$

kde Φ je funkce čtyř proměnných závisící též na funkci f . Podle toho, jak volíme funkci $\Phi(x_i, y_i, h_{i+1}, y_{i+1}; f)$, dostáváme různé metody.

U explicitní metody jsme po dosazení známé hodnoty y_0 do pravé strany (6.3) dostali hodnotu y_1 . Tu jsme opět dosadili do pravé strany (6.3) a dostali hodnotu y_2 atd.

U implicitních metod je situace odlišná. Po dosazení y_0 do pravé strany (6.5) dostaneme pro y_1 rovnici (obecně nelineární), protože y_1 se vyskytuje na obou stranách. Ta může být velmi složitá (záleží na složitosti funkcí Φ a f), takže se musí řešit numericky některou z metod kapitoly 2, nejčastěji Newtonovou. Podobně dosazením y_1 dostaneme rovnici pro y_2 atd.

Je přirozené ptát se, proč používáme implicitní metody, když vedou na mnohem složitější výpočty. Důvodem je, že mají obvykle daleko lepší vlastnosti, o nichž budeme mluvit dále (stabilita, rychlost konvergence).

Implicitní Eulerova metoda

Nejjednodušší implicitní jednokrokovou metodou je *implicitní Eulerova metoda*

$$y_{i+1} = y_i + h_{i+1}f(x_{i+1}, y_{i+1}), \quad i = 0, \dots, n-1. \quad (6.6)$$

V tomto případě je $\Phi(x_i, y_i, h_{i+1}, y_{i+1}; f) = f(x_i + h_{i+1}, y_{i+1})$. Tato metoda se někdy také nazývá zpětná Eulerova metoda.

Explicitní i implicitní metodu navrhl Euler již v roce 1768. Obě jsou předchůdkyněmi dnešních nesrovnatelně účinnějších metod.

Příklad 6.3 Napište vzorce Eulerovy explicitní a implicitní metody s konstantním krokem h pro počáteční úlohu $y' = x^2 + y^2$, $y(x_0) = y_0$.

Řešení. Protože $f(x, y) = x^2 + y^2$, dostaneme pro explicitní Eulerovu metodu ze vzorce (6.4)

$$y_{i+1} = y_i + h(x_i^2 + y_i^2), \quad i = 0, \dots, n-1.$$

Podobně pro implicitní Eulerovu metodu dostaneme ze vzorce (6.5)

$$y_{i+1} = y_i + h(x_{i+1}^2 + y_{i+1}^2), \quad i = 0, \dots, n-1.$$

Pro y_{i+1} tedy v tomto případě dostáváme kvadratickou rovnici

$$hy_{i+1}^2 - y_{i+1} + (y_i + hx_{i+1}^2) = 0.$$

Tu bychom mohli snadno vyřešit. V případě složitější pravé strany $f(x, y)$ by však bylo nutné řešit rovnici numericky. ▲

6.1.4 Vlastnosti jednokrokových metod

Než uvedeme další konkrétní jednokrokové metody, všimneme si, jaké vlastnosti by měly mít, aby bylo možné očekávat, že budou dávat dobré výsledky, tj. že numerická řešení jimi získaná se budou jen málo lišit od hodnot přesných řešení. Za tím účelem zavedeme několik důležitých pojmů. Protože explicitní metody (6.3) lze považovat za speciální případ implicitních metod (6.5) (Φ nezávisí na y_{i+1}), pokud to nebude nutné, budeme obecně mít na mysli druhý zmíněný vzorec.

Lokální a globální chyba metody, řád metody

Počáteční hodnota $y(x_0) = y_0$ je zadaná a je tudíž přesná. Hodnota y_1 , kterou z ní pomocí (6.3) nebo (6.5) vypočítáme, však už přesná nebude, platí jen $y(x_1) \approx y_1$. Z této přibližné hodnoty určíme y_2 , což bude také pouze přibližná hodnota pro $y(x_2)$. A tak dál. Obecně pro $i = 1, \dots, n$ platí $y(x_i) \approx y_i$. K výpočtu dosud neznámých hodnot tedy používáme nepřesné hodnoty, takže se (i při přesném počítání, kdybychom nějak vyloučili zaokrouhlovací chyby, což je ovšem nemožné) dopouštíme chyb. Tyto chyby se budou postupně kumulovat. Jde o typ chyby, který jsme v úvodu nazvali *chybou metody*.

Rozlišíme dvě chyby — tu, která vzniká v jednom kroku výpočtu, a tu, která vzniká kumulací chyb z jednotlivých kroků.

Uvažujme nějakou jedнокrokovou metodu.

- i) Předpokládejme, že hodnotu y_{i+1} na levé straně vzorce (6.5) vypočítáme bez zaokrouhlování z přesných hodnot $y(x_i)$ a $y(x_{i+1})$ (tzv. *lokalizační předpoklad*). Označme výslednou hodnotu \bar{y}_{i+1} . Výsledek nebude přesně $y(x_{i+1})$. Rozdíl

$$\begin{aligned} \text{lte}_i &= y(x_{i+1}) - \bar{y}_{i+1} = \\ &= y(x_{i+1}) - [y(x_i) + h_{i+1}\Phi(x_i, y(x_i), h_{i+1}, y(x_{i+1}); f)] \end{aligned}$$

nazýváme *lokální diskretizační chyba metody* (lte od local truncation error).

- ii) Ve skutečnosti ale hodnota y_i , $i \geq 1$, z níž počítáme y_{i+1} , není přesná. V i -tém kroku vlastně hledáme pomocí (6.5) hodnotu tzv. lokálního řešení $u_i(x)$ počáteční úlohy $u'_i = f(x, u_i)$, $u_i(x_i) = y_i$, v bodě x_{i+1} . Skutečná chyba, která vznikne, tj. rozdíl

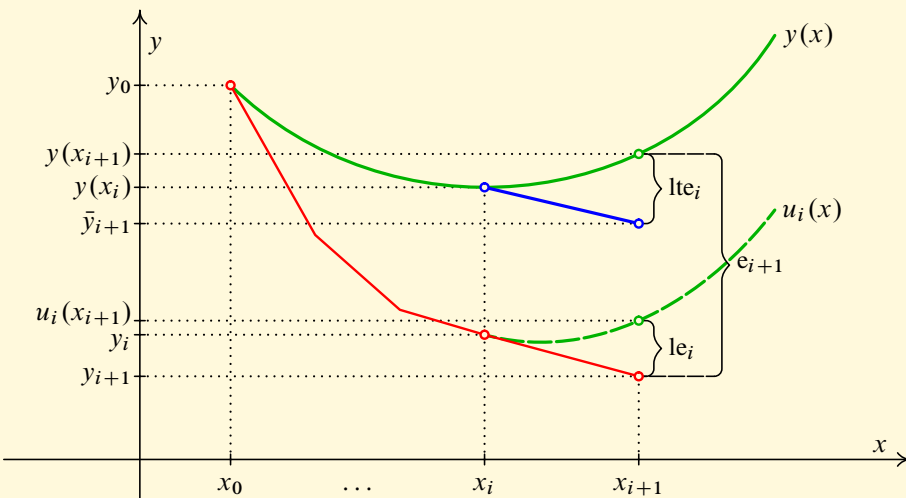
$$\begin{aligned} \text{le}_i &= u_i(x_{i+1}) - y_{i+1} = \\ &= u_i(x_{i+1}) - [y_i + h_{i+1}\Phi(x_i, y_i, h_{i+1}, y_{i+1}; f)], \end{aligned}$$

se nazývá *lokální chyba metody* (le od local error). Lze ukázat, že pro malé kroky jsou chyby l_{i-1} a l_i sice stejného řádu, ale stejné nejsou.

iii) Chyby se v jednotlivých krocích kumulují. Rozdíl mezi y_i a přesnou hodnotou $y(x_i)$, tj. $e_i = y(x_i) - y_i$, nazýváme *globální diskretizační chyba metody* po i krocích. Vztah mezi globální chybou a lokálními chybami je složitý. Rozhodně globální chyba není jejich prostým součtem. Lze jen říci, že $e_i \approx l_0 + \dots + l_{i-1}$. Dobrá metoda by měla dávat malé e_n .

Všechny tři druhy chyb jsou znázorněny na obr. 6.11.

V předcházejících definicích jsme uvažovali ideální situaci, kdy nedochází k zaokrouhlovacím chybám. To však v praxi není splněno, takže skutečná globální chyba je ovlivněna i tímto faktorem.



Obr. 6.11: Chyby jednokrokových metod

Uvažujme nějakou metodu a předpokládejme, že použijeme ekvidistantní dělení s krokem $h = \frac{b-a}{n}$. Největší přirozené číslo p , pro nějž platí (pro všechny počáteční úlohy s dostatečně hladkou pravou stranou), že $lte_i = O(h^{p+1})$, se nazývá *řád metody*. Tedy, aby byla konkrétní metoda řádu p , musí existovat konstanta C taková, že pro dostatečně malé kroky h platí $|lte_i| \leq C h^{p+1}$. Čím větší je tudíž p , tím menší je lokální diskretizační chyba (pro malá h se h^{p+1} zmenšuje s rostoucím p). Lze ukázat, že Eulerova explicitní i implicitní metoda je řádu jedna.

Následující úvaha naznačuje, co by mohlo platit pro přibližný odhad globální chyby, použijeme-li metodu řádu p . Protože $n = \frac{b-a}{h}$, je

$$\begin{aligned} e_n &\approx le_0 + \dots + le_{n-1} \approx lte_0 + \dots + lte_{n-1} = \underbrace{O(h^{p+1}) + \dots + O(h^{p+1})}_{n \text{ krát}} = \\ &= nO(h^{p+1}) = \frac{b-a}{h} O(h^{p+1}) = O(h^p). \end{aligned}$$

Vztah $e_n = O(h^p)$ skutečně za jistých předpokladů platí, důkaz je ale poměrně obtížný (přesná formulace a důkaz pro explicitní metody viz [52, str. 478]).

Konvergence metody

Důležitou otázkou je, zda se přibližné hodnoty při zmenšujícím se kroku přibližují k přesným hodnotám (předpokládáme, že počítáme bez zaokrouhlování).

Uvažujme nějakou jedнокrokovou metodu a zvolme ekvidistantní dělení s krokem $h = \frac{b-a}{n}$. Pro $n \rightarrow \infty$ bude platit $h \rightarrow 0$. Chtěli bychom, aby se přitom globální chyby e_i zmenšovaly. Zajímá nás proto veličina

$$\max\{|e_1|, \dots, |e_n|\} = \max\{|y(x_1) - y_1|, \dots, |y(x_n) - y_n|\}.$$

Řekneme, že metoda je *konvergentní*, jestliže pro všechny počáteční úlohy s dostatečně hladkou pravou stranou $f(x, y)$ platí

$$\lim_{h \rightarrow 0} \max\{|y(x_1) - y_1|, \dots, |y(x_n) - y_n|\} = 0.$$

Lze dokázat následující výsledek (viz [52, str. 478]):

Je-li jedнокroková metoda řádu alespoň jedna, pak je konvergentní.

Stabilita metody

Myšlenka, na níž je založen pojem, který si níže zavedeme, je velmi prostá. Budeme uvažovat jednoduchou testovací diferenciální rovnici s počáteční podmínkou

$$y' = \lambda y, \quad y(0) = 1, \quad (6.7)$$

kde λ je reálný nebo komplexní parametr, $\operatorname{Re} \lambda < 0$. Snadno najdeme její explicitní řešení $y(x) = e^{\lambda x}$. V případě, že $\lambda = \alpha + \beta i$, dostaneme s použitím Eulerova vztahu, že $y(x) = e^{\lambda x} = e^{\alpha x} \cos \beta x + i e^{\alpha x} \sin \beta x$, odkud $|y(x)| = \sqrt{e^{2\alpha x} \cos^2 \beta x + e^{2\alpha x} \sin^2 \beta x} = e^{\alpha x}$ (symbol i značí komplexní jednotku).

Protože je $\operatorname{Re} \lambda = \alpha < 0$, platí $\lim_{x \rightarrow \infty} y(x) = 0$. Od dobré numerické metody je přirozené očekávat, že při konstantním kroku $h > 0$ bude dávat řešení mající touž vlastnost, tedy

$$y_i \rightarrow 0 \quad \text{pro } x_i = ih \rightarrow \infty. \quad (6.8)$$

Označme Ω množinu všech komplexních čísel z tvaru $z = \lambda h$, kde pro dvojici λ, h dává daná numerická metoda pro testovací úlohu (6.7) řešení mající vlastnost (6.8).

(Jelikož předpokládáme, že $\operatorname{Re} \lambda < 0$ a $h > 0$, zajímají nás zejména čísla $z = \lambda h$, která leží vlevo od imaginární osy.)

Množina Ω se nazývá *oblast absolutní stability* uvažované numerické metody. Průnik množiny Ω se zápornou částí reálné osy se nazývá *interval absolutní stability*. Numerická metoda se nazývá *A-stabilní*, jestliže Ω obsahuje celou zápornou komplexní polorovinu $\mathbb{C}_- = \{\operatorname{Re} z < 0\}$.

Velikost oblasti absolutní stability dává omezení na délku kroku h , který může být použit při splnění vlastnosti (6.8). Čím menší tato množina je, tím musí být při daném λ krok h kratší. Nejlépe to demonstruje následující příklad.

Příklad 6.4 Určete oblast a interval absolutní stability explicitní a implicitní Eulerovy metody.

Řešení. Explicitní Eulerova metoda:

Ze vzorce (6.4) dostáváme pro počáteční úlohu (6.7)

$$y_{i+1} = y_i + h\lambda y_i = y_i(1 + \lambda h), \quad y_0 = 1,$$

takže

$$y_1 = 1 + \lambda h, \quad y_2 = (1 + \lambda h)^2, \dots, \quad y_i = (1 + \lambda h)^i.$$

Aby $y_i \rightarrow 0$ pro $i \rightarrow \infty$, musí platit $|1 + \lambda h| < 1$, tj. $|1 + z| < 1$, kde $z = \lambda h$. Oblast absolutní stability Ω je tedy otevřený jednotkový kruh se středem v bodě $[-1, 0]$ a interval absolutní stability je $(-2, 0)$. Tato metoda tudíž není A-stabilní (viz obr. 6.12 a)).

Implicitní Eulerova metoda:

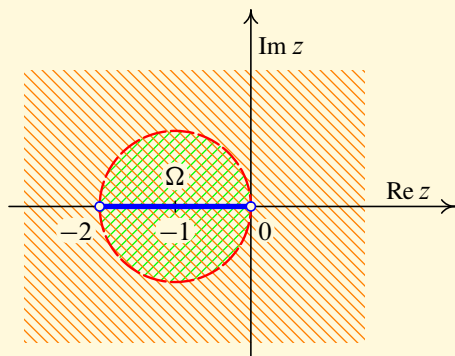
Ze vzorce (6.6) dostáváme pro počáteční úlohu (6.7)

$$y_{i+1} = y_i + h\lambda y_{i+1}, \quad y_0 = 1 \quad \Rightarrow \quad y_{i+1} = \frac{y_i}{1 - \lambda h},$$

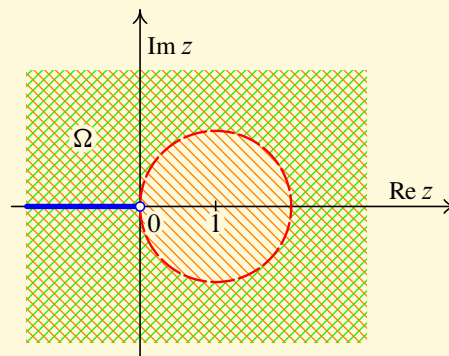
takže

$$y_1 = \frac{1}{1 - \lambda h}, \quad y_2 = \frac{1}{(1 - \lambda h)^2}, \dots, \quad y_i = \frac{1}{(1 - \lambda h)^i}.$$

Aby $y_i \rightarrow 0$ pro $i \rightarrow \infty$, musí platit $1/|1 - \lambda h| < 1$, tj. $|1 - z| > 1$, kde $z = \lambda h$. Oblast absolutní stability Ω je proto vnějšek otevřeného jednotkového kruhu se středem v bodě $[1, 0]$ a obsahuje zřejmě celou zápornou komplexní polorovinu \mathbb{C}_- . Interval absolutní stability je $(-\infty, 0)$. Tato metoda je tudíž A-stabilní (viz obr. 6.12 b)).



a) Explicitní Eulerova metoda



b) Implicitní Eulerova metoda

Obr. 6.12: Oblast a interval absolutní stability

Zvolme např. $\lambda = -50$. Řešení $y(x) = e^{-50x}$ počáteční úlohy (6.7) se s rostoucím x velice rychle zmenšuje a brzy je téměř nulové. Abychom pomocí explicitní Eulerovy metody dostali hodnoty y_i , pro něž $y_i \rightarrow 0$ pro $i \rightarrow \infty$, musíme krok h volit tak, aby číslo $-50h$ leželo v intervalu $(-2, 0)$, tj. $h < 0,04$. Tedy přes skutečnost, že přesné řešení je téměř nulové, musíme krok volit velice malý.

Oproti tomu u implicitní Eulerovy metody lze krok h volit zcela libovolně. ▲

Situace popsaná v předchozím příkladu je typická nejen pro zmíněné Eulerovy metody. Implicitní jednokrokové metody mají obecně daleko větší oblast absolutní stability, takže je možné volit mnohem delší kroky. Tím se snižuje počet aritmetických operací potřebných k nalezení hodnoty $y_n \approx y(b)$. Přestože je nutné v každém kroku implicitní metody řešit obecně nelineární rovnici, je jejich použití mnohdy daleko účinnější.

Doposud jsme předpokládali, že počítáme bez zaokrouhlovacích chyb, a věnovali jsme se jen chybám metod. V praxi však k zaokrouhlování dochází. Obecně při zmenšování kroku h se zmenšuje globální chyba metody, ale zvětšuje se zaokrouhlovací chyba (což je pochopitelné, protože se provádí více aritmetických výpočtů). Celková chyba je pak jejich součtem. Je proto potřeba najít vhodný kompromis, aby některá z chyb výrazně nepřevládla a neznehodnotila zásadně výsledek.

6.1.5 Metody Rungeho-Kutty

Důležitou skupinu jednokrokových metod navrhli Runge¹ a Kutta². Při výpočtu hodnoty y_{i+1} se vypočítají hodnoty směřového pole, tj. funkce $f(x, y)$, v několika vhodně zvolených pomocných bodech, které jsou rozmístěny kolem bodu $[x_i, y_i]$. Počet pomocných bodů udává tzv. stupeň dané metody. Tyto hodnoty se pak zprůměrují (udělá se jejich vážený průměr) a výsledek se použije jako směrnice $\Phi(x_i, y_i, h_{i+1}; f)$ ve vzorci (6.3) resp. $\Phi(x_i, y_i, h_{i+1}, y_{i+1}; f)$ ve vzorci (6.5).

Obecný tvar *s-stupňové metody Rungeho-Kutty*, kde $s \in \mathbb{N}$, $s \geq 1$, s konstantním krokem h vypadá takto:

$$y_{i+1} = y_i + h(\gamma_1 k_1 + \gamma_2 k_2 + \cdots + \gamma_s k_s), \quad i = 0, \dots, n-1, \quad (6.9)$$

¹**Carl David Tolmé Runge** (1856–1927) — německý matematik, známý aplikacemi matematiky ve fyzice.

²**Martin Wilhelm Kutta** (1867–1944) — německý matematik a inženýr, známý numerickou metodou řešení obyčejných diferenciálních rovnic.

kde

$$\begin{aligned} k_1 &= f(x_i + \alpha_1 h, y_i + h(\beta_{11}k_1 + \beta_{12}k_2 + \cdots + \beta_{1s}k_s)), \\ k_2 &= f(x_i + \alpha_2 h, y_i + h(\beta_{21}k_1 + \beta_{22}k_2 + \cdots + \beta_{2s}k_s)), \\ k_3 &= f(x_i + \alpha_3 h, y_i + h(\beta_{31}k_1 + \beta_{32}k_2 + \cdots + \beta_{3s}k_s)), \\ &\vdots \\ k_s &= f(x_i + \alpha_s h, y_i + h(\beta_{s1}k_1 + \beta_{s2}k_2 + \cdots + \beta_{ss}k_s)) \end{aligned}$$

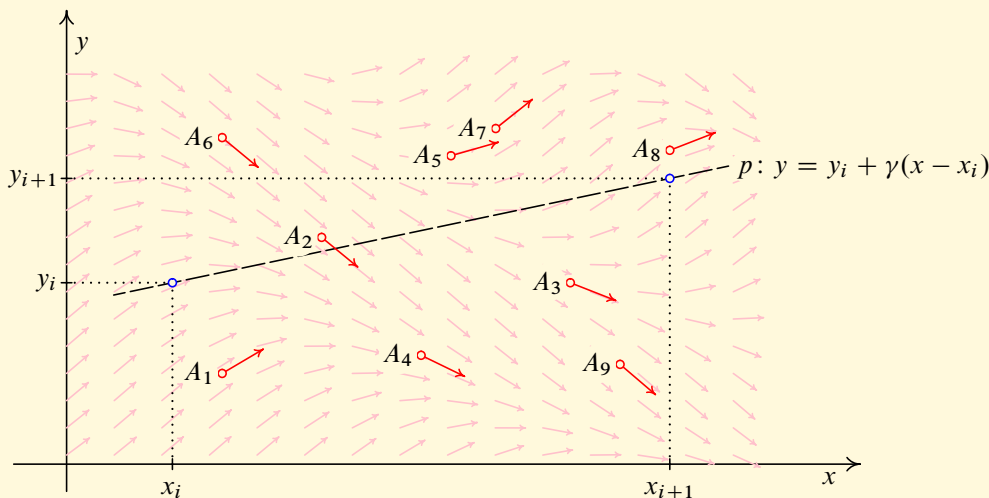
a $\alpha_r, \beta_{rt}, \gamma_t, r, t = 1, \dots, s$, jsou nějaké reálné konstanty. Označíme-li

$$X_r = x_i + \alpha_r h \quad \text{a} \quad Y_r = y_i + h(\beta_{r1}k_1 + \beta_{r2}k_2 + \cdots + \beta_{rs}k_s), \quad r = 1, \dots, s,$$

bude mít vzorec (6.9) tvar

$$y_{i+1} = y_i + h(\gamma_1 f(X_1, Y_1) + \gamma_2 f(X_2, Y_2) + \cdots + \gamma_s f(X_s, Y_s)), \quad i = 0, \dots, n-1.$$

Jde tedy o vzorec typu (6.3) resp. (6.5), kde funkce Φ , mající geometrický význam směrnice přímky procházející bodem $[x_i, y_i]$, je váženým průměrem hodnot směrového pole v s pomocných bodech $[X_1, Y_1], \dots, [X_s, Y_s]$ s vahami $\gamma_1, \dots, \gamma_s$, viz obr. 6.13.



Obr. 6.13: Princip metod Rungeho-Kutty

$s = 9$ (devítistupňová metoda)

$A_i = [X_i, Y_i], i = 1 \dots, 9$ (pomocné body)

$\gamma = \gamma_1 f(X_1, Y_1) + \dots + \gamma_9 f(X_9, Y_9)$ (směrnice přímky p)

Abychom dostali konkrétní metodu, je třeba vhodně určit koeficienty α_r , β_{rt} a γ_t , $r, t = 1, \dots, s$. Snahou je dosáhnout toho, aby metoda měla co nejvyšší řád. Postup je založen na Taylorově rozvoji lokální diskretizační chyby. Odvození je technicky velice komplikované (viz např. [7, str. 150], [9, str. 17], [23], [37, str. 165] nebo [43, str. 219]). Obdržíme soustavu rovnic pro neznámé koeficienty. Ty nejsou určeny jednoznačně, zůstanou volné parametry, které se volí tak, aby metoda měla další dobré vlastnosti (stabilita, vyladění chyby apod.). V praxi se používají metody pro stupeň s roven dva až třináct.

Koeficienty se zapisují do tzv. *Butcherovy*¹ tabulky:

α_1	β_{11}	β_{12}	\dots	β_{1s}
α_2	β_{21}	β_{22}	\dots	β_{2s}
\vdots	\vdots	\vdots	\ddots	\vdots
α_s	β_{s1}	β_{s2}	\dots	β_{ss}
	γ_1	γ_2	\dots	γ_s

¹**John Charles Butcher** (1933) (čti bučr) — novozélandský matematik, zabývá se numerickým řešením obyčejných diferenciálních rovnic.

Obecně platí, že $\gamma_1 + \gamma_2 + \dots + \gamma_s = 1$. Čísla γ_t mohou být nulová i záporná. Navíc pro všechny prakticky používané metody platí

$$\alpha_r = \beta_{r1} + \dots + \beta_{rs}, \quad r = 1, \dots, s. \quad (6.10)$$

V dalším budeme předpokládat, že je tato vlastnost splněna.

Metoda je *explicitní*, pokud má matice (β_{rt}) nenulové prvky pouze pod hlavní diagonálou. Hodnoty k_1, \dots, k_s se postupně snadno vypočítají. S ohledem na zmíněný předpoklad totiž platí:

$$\alpha_1 = 0 + 0 + \dots + 0 + 0 = 0,$$

$$\alpha_2 = \beta_{21} + 0 + \dots + 0 + 0 = \beta_{21},$$

$$\alpha_3 = \beta_{31} + \beta_{32} + \dots + 0 + 0 = \beta_{31} + \beta_{32},$$

$$\vdots$$

$$\alpha_s = \beta_{s1} + \beta_{s2} + \dots + \beta_{ss-1} + 0 = \beta_{s1} + \beta_{s2} + \dots + \beta_{ss-1}.$$

Pro k_1, \dots, k_s tak v případě explicitní metody dostaneme, že

$$k_1 = f(x_i, y_i),$$

$$k_2 = f(x_i + \alpha_2 h, y_i + h\beta_{21}k_1),$$

$$k_3 = f(x_i + \alpha_3 h, y_i + h(\beta_{31}k_1 + \beta_{32}k_2)),$$

$$\vdots$$

$$k_s = f(x_i + \alpha_s h, y_i + h(\beta_{s1}k_1 + \beta_{s2}k_2 + \dots + \beta_{s,s-1}k_{s-1})).$$

V opačném případě se jedná o *implicitní* metodu. Pak máme pro k_1, \dots, k_s soustavu obecně nelineárních rovnic, která se musí řešit numericky. Lze ukázat, že pro malé kroky h má soustava vždy jediné řešení, pokud f splňuje Lipschitzovu podmínku (viz [7, str. 214]).

Metoda se nazývá *semiimplicitní*, je-li implicitní a matice (β_{rt}) má nad hlavní diagonálou nulové prvky. V tomto případě se z první rovnice vypočítá k_1 a dosadí se do druhé rovnice, z té se vypočítá k_2 atd. Stačí tedy numericky řešit pouze rovnice o jedné neznámé.

Pro daný počet stupňů s označme $p(s)$ maximální řád s -stupňové metody, který lze dosáhnout. Pro explicitní metody je o funkci $p(s)$ známo, že $p(s) \rightarrow \infty$ pro $s \rightarrow \infty$

a dále:

$$\begin{aligned} p(s) &= s & \text{pro } s &= 1, 2, 3, 4, & p(s) &= s - 2 & \text{pro } s &= 8, 9, \\ p(s) &= s - 1 & \text{pro } s &= 5, 6, 7, & p(s) &\leq s - 2 & \text{pro } s &\geq 10. \end{aligned}$$

Situace je jiná u implicitních metod, u nichž může být $p(s) > s$.

Pokud jde o konvergenci a její rychlost, lze dokázat, že metoda Rungeho-Kutty řádu p má pro diferenciální rovnici s dostatečně hladkou pravou stranou $f(x, y)$ globální diskretizační chybu $e_i = y(x_i) - y_i = O(h^p)$, viz [9, str. 20] nebo [57, str. 41].

Dále si všimneme stability těchto metod. Použijeme-li takovou metodu na testovací úlohu (6.7), vyjde nám, že $y_{i+1} = R(z)y_i$, tj. $y_i = [R(z)]^i$, $i = 1, 2, \dots$, kde $z = \lambda h$. Funkce $R(z)$ se nazývá *funkce stability* dané metody. Pro oblast absolutní stability tudíž dostaneme, že $\Omega = \{z \in \mathbb{C} : |R(z)| < 1\}$. U explicitních s -stupňových metod je $R(z)$ polynomem stupně s , takže tyto metody nikdy nemohou být A-stabilní. U implicitních s -stupňových metod je $R(z)$ racionální funkce, jejíž čitatel a jmenovatel jsou polynomy stupně nejvýše s , viz [7, str. 230]. V příkladu 6.4 jsme určili, že u explicitní Eulerovy metody je $R(z) = 1 + z$ a u implicitní Eulerovy metody je $R(z) = \frac{1}{1-z}$.

U některých typů rovnic ani A-stabilní metody nedávají dobré výsledky. Proto se zavádí ještě silnější pojem stability. Metoda typu Rungeho-Kutty se nazývá *L-stabilní*, jestliže je A-stabilní a navíc platí, že $|R(z)| \rightarrow 0$ pro $\operatorname{Re} z \rightarrow -\infty$. Tedy např. Eulerova implicitní metoda je L-stabilní. Viz [57, str. 116].

Existuje celá řada metod typu Rungeho-Kutty. Zmíníme pouze několik klasických příkladů explicitních metod. Uvedeme vždy vzorce a odpovídající Butcherovu tabulku. Doplňme, že jediná jednostupňová explicitní metoda je Eulerova metoda (6.4).

1) *První modifikace Eulerovy metody* (Collatz¹ 1960):

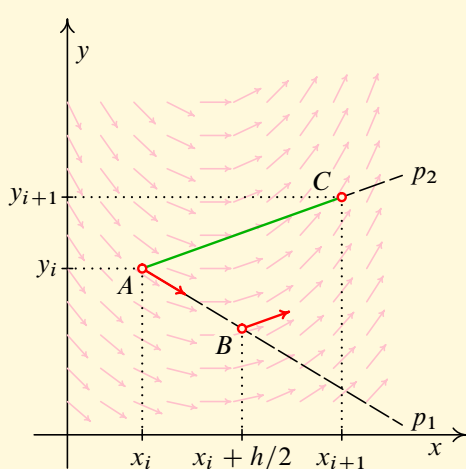
$$\begin{aligned} y_{i+1} &= y_i + hk_2, & i &= 0, \dots, n-1, \\ k_1 &= f(x_i, y_i), \\ k_2 &= f(x_i + h/2, y_i + hk_1/2). \end{aligned}$$

0	0	0
1/2	1/2	0
	0	1

Je to dvoustupňová metoda druhého řádu. Princip vyčíslení viz obr. 6.14 a) (význam bodů *A*, *B* a *C* je podobný jako na obr. 6.15). Funkce Φ ze vztahu (6.3) má tvar

$$\Phi(x_i, y_i, h; f) = f(x_i + h/2, y_i + hf(x_i, y_i)/2).$$

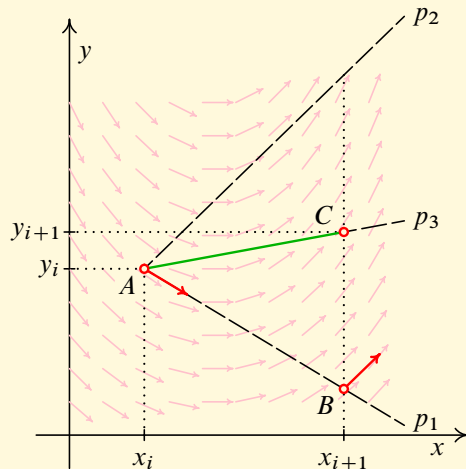
¹Lothar Collatz (1910–1990) (čti kolac) — německý matematik. Zabýval se numerickou matematikou.



a) První modifikace Eulerovy metody

$$p_1: y = y_i + k_1(x - x_i)$$

$$p_2: y = y_i + k_2(x - x_i)$$



b) Druhá modifikace Eulerovy metody

$$p_1: y = y_i + k_1(x - x_i)$$

$$p_2: y = y_i + k_2(x - x_i)$$

$$p_3: y = y_i + \frac{k_1 + k_2}{2}(x - x_i)$$

Obr. 6.14: Dvoustupňové metody typu Rungeho-Kutty

2) *Druhá modifikace Eulerovy metody* též zvaná *Heunova¹ metoda* (Heun 1900):

$$y_{i+1} = y_i + h \frac{k_1 + k_2}{2}, \quad i = 0, \dots, n-1,$$

$$k_1 = f(x_i, y_i),$$

$$k_2 = f(x_i + h, y_i + hk_1).$$

0	0	0
1	1	0
	1/2	1/2

Je to dvoustupňová metoda druhého řádu. Princip vyčíslení viz obr. 6.14 b) (význam bodů A , B a C je podobný jako na obr. 6.15). Funkce Φ ze vztahu (6.3) má tvar

$$\Phi(x_i, y_i, h; f) = [f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))]/2.$$

Podobně lze tuto funkci snadno vypsát pro další metody, takže příslušné vzorce nebudeme uvádět.

¹**Karl Heun** (1859–1929) (čti hojn) — německý matematik.

3) *Kuttova metoda* (Kutta 1901):

$$y_{i+1} = y_i + h \frac{k_1 + 4k_2 + k_3}{6},$$

$$i = 0, \dots, n-1,$$

$$k_1 = f(x_i, y_i),$$

$$k_2 = f(x_i + h/2, y_i + hk_1/2),$$

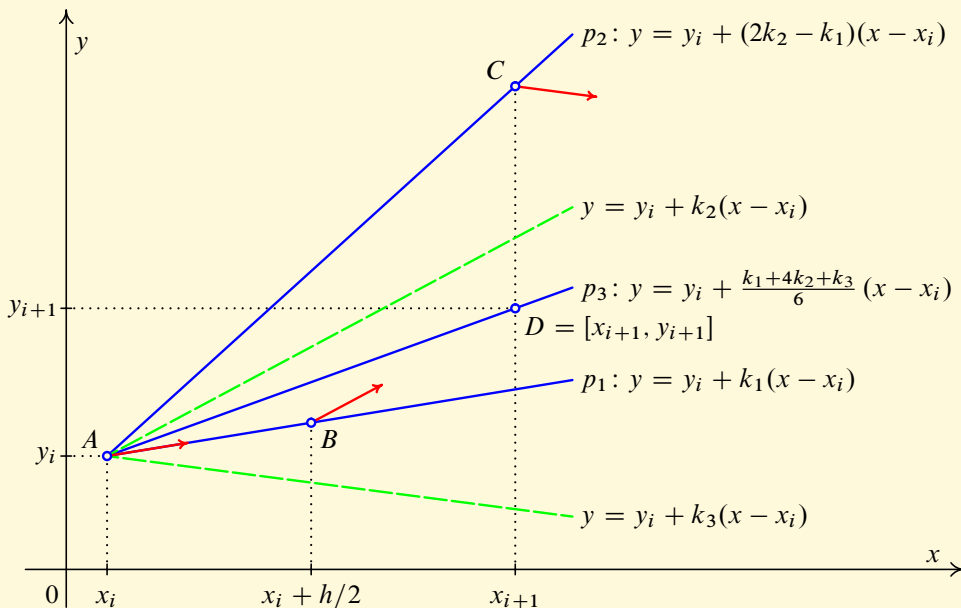
$$k_3 = f(x_i + h, y_i + h(2k_2 - k_1)).$$

0	0	0	0
1/2	1/2	0	0
1	-1	2	0
	1/6	4/6	1/6

Je to třístupňová metoda třetího řádu. Ukážeme si na ní podrobněji geometrický princip rozmístění pomocných bodů — viz obr. 6.15.

- Označme k_1 směrnici vektoru směřového pole v bodě $A = [x_i, y_i]$. Sestrojíme přímku p_1 mající směrnici k_1 , která prochází bodem A .
Tedy $p_1: y = y_i + k_1(x - x_i)$.
- Na přímce p_1 najdeme bod B , jehož x -ová souřadnice je $x_i + h/2$. Jeho souřadnice jsou $B = [x_i + h/2, y_i + hk_1/2]$.

- Označme k_2 směrnici vektoru směřového pole v bodě B . Sestrojíme přímku p_2 mající směrnici $2k_2 - k_1$, která prochází bodem A .
Tedy $p_2: y = y_i + (2k_2 - k_1)(x - x_i)$.
- Na přímce p_2 najdeme bod C , jehož x -ová souřadnice je $x_i + h$. Jeho souřadnice jsou $C = [x_i + h, y_i + h(2k_2 - k_1)]$.
- Označme k_3 směrnici vektoru směřového pole v bodě C . Sestrojíme přímku p_3 mající směrnici $(k_1 + 4k_2 + k_3)/6$, která prochází bodem A .
Tedy $p_3: y = y_i + (k_1 + 4k_2 + k_3)/6 \cdot (x - x_i)$.
- Na přímce p_3 najdeme bod D , jehož x -ová souřadnice je $x_i + h$. Jeho souřadnice jsou $D = [x_i + h, y_i + h(k_1 + 4k_2 + k_3)/6]$. Nyní y -ová souřadnice bodu D je y_{i+1} .



Obr. 6.15: Kuttova metoda

4) *Klasická metoda Rungeho-Kutty* (Runge 1895, Kutta 1901):

$$y_{i+1} = y_i + h \frac{k_1 + 2k_2 + 2k_3 + k_4}{6},$$

$$i = 0, \dots, n-1,$$

$$k_1 = f(x_i, y_i),$$

$$k_2 = f(x_i + h/2, y_i + hk_1/2),$$

$$k_3 = f(x_i + h/2, y_i + hk_2/2),$$

$$k_4 = f(x_i + h, y_i + hk_3).$$

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	2/6	2/6	1/6

Je to čtyřstupňová metoda čtvrtého řádu. Tato metoda byla pro svou jednoduchost oblíbená v éře před nástupem počítačů.

5) *Gillova metoda* (Gill 1951):

$$y_{i+1} = y_i + \frac{h}{6} \left[k_1 + (2 - \sqrt{2})k_2 + (2 + \sqrt{2})k_3 + k_4 \right], \quad i = 0, \dots, n-1,$$

$$k_1 = f(x_i, y_i),$$

$$k_2 = f(x_i + h/2, y_i + hk_1/2),$$

$$k_3 = f(x_i + h/2, y_i + h((-1/2 + 1/\sqrt{2})k_1 + (1 - 1/\sqrt{2})k_2)),$$

$$k_4 = f(x_i + h, y_i + h(-k_2/\sqrt{2} + (1 + 1/\sqrt{2})k_3)).$$

0	0	0	0	0
1/2	1/2	0	0	0
1/2	$-1/2 + 1/\sqrt{2}$	$1 - 1/\sqrt{2}$	0	0
1	0	$-1/\sqrt{2}$	$1 + 1/\sqrt{2}$	0
	1/6	$(2 - \sqrt{2})/6$	$(2 + \sqrt{2})/6$	1/6

Je to čtyřstupňová metoda čtvrtého řádu. U této metody jsou redukovány paměťové nároky, viz [7, str. 180].

Řadu dalších explicitních i implicitních metod typu Rungeho-Kutty lze nalézt např. v [7], [9], [37] nebo [72].

Současné metody

Je třeba zdůraznit, že soudobé profesionální programy na numerické řešení obyčejných diferenciálních rovnic nepoužívají přímo výše uvedené tzv. klasické jedнокrokové metody, ale daleko efektivnější metody, které jsou na klasických metodách založené. Jde o velmi účinné tzv. *adaptivní metody*, u nichž se provádí pomocí důmyslných algoritmů řízení délky kroku. Základní myšlenkou je udržet odhad velikosti lokální chyby le_i zhruba stejně veliký. Tomu se přizpůsobuje délka kroku. Pro její řízení se používají dvě strategie — EPS (error per step) a EPUS (error per unit step). K rozhodnutí o délce kroku je nutný odhad velikosti lokální chyby. Pro něj se používá např. metoda polovičního kroku a dnes zejména metoda vnořených odhadů chyby. Metody založené na posledně zmíněném principu se nazývají *vnořené (embedded) metody Rungeho-Kutty*. Jsou tvořeny dvojicí metod Rungeho-Kutty řádu p a $p + 1$. Pokud se ve výpočtu pokračuje přesnější metodou, říká se, že byla použita dvojice metod s lokální extrapolací. Když se pokračuje méně

přesnou metodou a přesnější metoda byla použita jen k odhadu lokální chyby, říká se, že dvojice metod byla použita bez lokální extrapolace. Více podrobností lze nalézt například v [9, str. 23–33] nebo [11, str. 15–18]. Důkladný teoretický výklad je uveden v [7], [37] a [50]. Na stránce [67] lze kromě přehledu najít zdrojové kódy v C nejnámějších metod tohoto typu. Jejich dobrá implementace totiž rozhodně není triviální záležitostí. Stručný přehled historie odvozování vhodných dvojic explicitních metod lze najít v [76], ukázky odvození v [74].

V programu Maple můžeme najít (kromě jiných) následující jednokrokové metody:

- rkf45 — dvojice šestistupňových metod řádu 4 a 5 (autor Erwin Fehlberg, 1968),
- ck45 — dvojice šestistupňových metod řádu 4 a 5 (autoři Jeff. R. Cash a Alan H. Karp, 1988),
- dverk78 — dvojice třináctistupňových metod řádu 7 a 8 (autor Jim Verner, [77]).

6.1.6 Vícekrokové metody

Nyní si všimneme vícekrokových metod pro řešení počáteční úlohy $y' = f(x, y)$, $y(x_0) = y_0$, na intervalu $\langle a, b \rangle$, $a = x_0$. Pro lepší přehlednost označíme $f(x_i, y_i) = f_i$. Budeme předpokládat, že je použit konstantní krok $h > 0$. Omezíme se jen na tzv. *lineární vícekrokové metody* (stručně LVM; název lineární je použit proto, protože vzorce těchto metod budou na symbolech y_j a f_j záviset lineárně).

Obecná lineární k -kroková metoda, kde $k \in \mathbb{N}$, má tvar

$$\sum_{j=0}^k a_j y_{i+1-j} = h \sum_{j=0}^k b_j f_{i+1-j}, \quad (6.11)$$

tj. po rozepsání sum

$$\begin{aligned} a_0 y_{i+1} + a_1 y_i + a_2 y_{i-1} + \cdots + a_k y_{i+1-k} &= \\ &= h(b_0 f_{i+1} + b_1 f_i + b_2 f_{i-1} + \cdots + b_k f_{i+1-k}), \end{aligned}$$

kde a_j, b_j jsou vhodné konstanty. Předpokládáme, že $a_0 \neq 0$ a $a_k^2 + b_k^2 > 0$. Tedy bez újmy na obecnosti můžeme předpokládat, že metoda je *normalizovaná*, tj. $a_0 = 1$

(původní rovnici vydělíme a_0), a osamostatníme y_{i+1} . Dostaneme:

$$\begin{aligned} y_{i+1} &= - \sum_{j=1}^k a_j y_{i+1-j} + h \sum_{j=0}^k b_j f_{i+1-j} = & (6.12) \\ &= -a_1 y_i - a_2 y_{i-1} - \cdots - a_k y_{i+1-k} + \\ &\quad + h(b_0 f_{i+1} + b_1 f_i + \cdots + b_k f_{i+1-k}). \end{aligned}$$

Jak již bylo dříve zmíněno, aby se mohla taková metoda (pro $k > 1$) použít, je třeba k zadané hodnotě y_0 nějak učit (vhodnou jedнокrokovou metodou) co nejpřesněji startovací hodnoty y_1, y_2, \dots, y_{k-1} .

Pokud je $b_0 = 0$, vyskytuje se y_{i+1} jen na levé straně (6.11) a lze ho osamostatnit, viz (6.12). Takové LVM se nazývají *explicitní*. Je-li však $b_0 \neq 0$, dostáváme pro y_{i+1} obecně nelineární rovnici tvaru

$$y_{i+1} = A + h b_0 f(x_{i+1}, y_{i+1}),$$

kde

$$A = -a_1 y_i - a_2 y_{i-1} - \cdots - a_k y_{i+1-k} + h(b_1 f_i + \cdots + b_k f_{i+1-k}),$$

která se musí řešit numericky. Takové LVM se nazývají *implicitní*. Jejich předností je, že mají lepší vlastnosti; viz následující oddíl.

K nejznámějším LVM patří *Adamsovy metody* a *metody zpětného derivování*. Než si však všimneme konkrétních vzorců, podíváme se na vlastnosti, které by dobré numerické metody tohoto typu měly mít.

6.1.7 Vlastnosti vícezkrokových metod

Zavedeme obdobné pojmy jako v oddílu 6.1.4 pro jednokrokové metody, abychom mohli popsat vlastnosti dobrých LVM.

Lokální a globální chyba metody, řád metody

Do levé a pravé strany vzorce (6.11) dosadíme hodnoty přesného řešení (tzv. *lokalizační předpoklad*). Výsledky budou obecně různé. Jejich rozdíl

$$l_{te_i} = \sum_{j=0}^k a_j y(x_{i+1-j}) - h \sum_{j=0}^k b_j f(x_{i+1-j}, y(x_{i+1-j}))$$

nazveme *lokální diskretizační chyba metody*. Na rozdíl od jednokrokových metod nebudeme zavádět pojem (skutečné) lokální chyby, protože nemá praktický význam.

Řekneme, že daná *metoda je řádu* p , $p \in \mathbb{N}$, jestliže p je největší číslo, pro něž platí $l_{te_i} = O(h^{p+1})$.

Rozdíl $e_i = y(x_i) - y_i$ nazveme *globální diskretizační chyba metody* po i krocích.

Při popisu vlastností LVM hrají důležitou roli dva polynomy, jejichž koeficienty jsou obsaženy v (6.11):

$$\begin{aligned}\rho(\lambda) &= a_0\lambda^k + a_1\lambda^{k-1} + \dots + a_{k-1}\lambda + a_k, \\ \sigma(\lambda) &= b_0\lambda^k + b_1\lambda^{k-1} + \dots + b_{k-1}\lambda + b_k.\end{aligned}$$

Nazývají se *první a druhý charakteristický polynom*.

Metoda se nazývá *konzistentní*, jestliže je řádu alespoň jedna. Lze dokázat následující tvrzení: *LVM je konzistentní právě tehdy, když $\rho(1) = 0$ a $\rho'(1) = \sigma(1)$* (viz [9, str. 37], [19, str. 370], [52, str. 510] nebo [57, str. 67]).

D-stabilita a konvergence metody

Při zavádění následujícího pojmu opět vyjdeme z vhodné testovací úlohy podobně jako při definici A-stability, tentokrát dokonce ještě jednodušší. Půjde o počáteční úlohu

$$y' = 0, \quad y(0) = 0 \quad \text{pro } x \in \langle 0, 1 \rangle.$$

Přesným řešením je pochopitelně funkce $y(x) = 0$. Ze vztahu (6.11) dostaneme rekurentní formuli

$$a_0 y_{i+1} + a_1 y_i + a_2 y_{i-1} + \cdots + a_k y_{i+1-k} = 0. \quad (6.13)$$

Rozdělíme-li interval $\langle 0, 1 \rangle$ ekvidistantně na n dílků, tj. $h = 1/n$, a zvolíme-li $y_1 = y_2 = \cdots = y_{k-1} = 0$, vyjde postupně $y_k = y_{k+1} = \cdots = y_n = 0$. Nás však bude zajímat co se stane, když hodnoty y_1, y_2, \dots, y_{k-1} budou mírně porušené (obecně jsou určovány nějakou vhodnou jedнокrokovou metodou a nejsou tudíž přesné). Je přirozené požadovat, aby hodnota y_n aproximující přesnou hodnotu $y(1) = 0$ byla s rostoucím n malá, pokud je počítaná z výchozích hodnot y_1, y_2, \dots, y_{k-1} , které se jen málo liší od nuly.

Rovnice (6.13) se nazývá *lineární diferenční rovnice s konstantními koeficienty*. Teorie těchto rovnic je velmi podobná teorii lineárních diferenciálních rovnic s konstantními koeficienty, viz např. [52, str. 501] nebo [68, str. 20]. Ukazuje se, že výše zmíněný požadavek platí právě tehdy, když kořeny prvního charakteristického polynomu leží v komplexní rovině v jednotkovém kruhu se středem v počátku, přičemž ty, které jsou na hranici, jsou jednoduché. Z tohoto důvodu se zavádí následující definice.

Definice 6.5 LVM se nazývá *D-stabilní* neboli *stabilní ve smyslu Dahlquista*¹, jestliže pro všechny kořeny ζ prvního charakteristického polynomu $\rho(\lambda)$ platí $|\zeta| \leq 1$, přičemž pokud je $|\zeta| = 1$, je tento kořen jednoduchý.

¹**Germund Dahlquist** (1925–2005) (čti dalkvist) — švédský matematik, patřil k průkopníkům zkoumání problematiky stability numerického řešení diferenciálních rovnic.

Je známo, že pro řád p D-stabilních k -krokových metod platí:

$$\begin{aligned} p &\leq k && \text{pro explicitní metody,} \\ p &\leq k + 1 && \text{pro implicitní metody a } k \text{ liché,} \\ p &\leq k + 2 && \text{pro implicitní metody a } k \text{ sudé} \end{aligned}$$

(tzv. **první Dahlquistova bariéra**). Viz [19, str. 384], [57, str. 90].

Definice konvergence LVM je obdobná jako definice jednokrokových metod na str. 600. Lineární k -kroková metoda se nazývá *konvergentní*, jestliže pro všechny počáteční úlohy s dostatečně hladkou pravou stranou $f(x, y)$ platí

$$\lim_{h \rightarrow 0} \max\{|y(x_1) - y_1|, \dots, |y(x_n) - y_n|\} = 0, \quad h = (b - a)/n,$$

přičemž tato vlastnost má platit, pokud počáteční hodnoty y_1, y_2, \dots, y_{k-1} (které nemusí být přesné), z nichž se počítají další hodnoty y_k, \dots, y_n , se s rostoucím n přibližují k přesným hodnotám $y(x_1), y(x_2), \dots, y(x_{k-1})$.

Zatímco jednokrokové metody řádu aspoň jedna byly vždy konvergentní, u vícekových metod je situace složitější. Lze dokázat následující výsledek (viz např. [52, str. 506], [57, str. 65,72] nebo [68, str. 31]): *LVM je konvergentní právě tehdy, když je konzistentní a D-stabilní.*

O rychlosti konvergence lze dokázat následující výsledek (viz např. [52, str. 508] nebo [57, str. 75]):

Je-li konvergentní LVM řádu $p \geq 1$, jsou-li počáteční hodnoty y_1, y_2, \dots, y_{k-1} zadány s chybou $O(h^p)$ a pravá strana $f(x, y)$ diferenciální rovnice má spojitě derivace až do řádu p , platí pro globální diskretizační chybu $e_i = y(x_i) - y_i = O(h^p)$.

A-stabilita

Vrátíme se opět k testovací úloze (6.7). Po dosazení do (6.11) dostaneme

$$a_0 y_{i+1} + a_1 y_i + \dots + a_k y_{i+1-k} = h(b_0 \lambda y_{i+1} + b_1 \lambda y_i + \dots + b_k \lambda y_{i+1-k}),$$

odkud po úpravě vyjde

$$(a_0 - h\lambda b_0) y_{i+1} + (a_1 - h\lambda b_1) y_i + \dots + (a_k - h\lambda b_k) y_{i+1-k} = 0.$$

Polynom

$$\pi(\mu, z) = \sum_{j=0}^k (a_j - zb_j)\mu^{k-j} = \rho(\mu) - z\sigma(\mu),$$

proměnné μ , kde $z = \lambda h$, nazýváme *polynom stability* LVM. Koeficienty tohoto polynomu tedy závisí na komplexním parametru z . Připomeňme, že $\rho(\mu)$ a $\sigma(\mu)$ jsou první a druhý charakteristický polynom (viz str. 627).

Vztah (6.8) platí právě tehdy, když pro všechny kořeny ζ polynomu stability $\pi(\mu, z)$ (μ je proměnná, z je parametr) platí $|\zeta| < 1$. Označme Ω množinu všech těch komplexních čísel z , pro něž kořeny polynomu $\pi(\mu, z)$ mají tuto vlastnost. Množina Ω se nazývá *oblast absolutní stability* LVM. Tato množina je důležitou charakteristikou LVM. Metody s velkou oblastí absolutní stability jsou vhodné pro řešení tzv. tuhých problémů (viz oddíl 6.1.9). Obsahuje-li Ω celou zápornou komplexní polorovinu, nazývá se metoda *A-stabilní*. Je známo, že explicitní LVM nemohou být A-stabilní a implicitní LVM, které jsou A-stabilní, mohou mít nejvýše řád $p = 2$ (tzv. **druhá Dahlquistova bariéra**). Viz [57, str. 113].

6.1.8 Příklady lineárních více krokových metod

Uvedeme přehled nejvýznamnějších LVM. Podrobnosti o jejich odvození, odhadech lokálních chyb a oblastech absolutní stability lze nalézt např. v [7, 9, 11, 37, 52, 68].

Adamsovy metody

Integrací rovnice (6.2) na intervalu $\langle x_i, x_{i+1} \rangle$ dostaneme vztah

$$y(x_{i+1}) - y(x_i) = \int_{x_i}^{x_{i+1}} f(x, y(x)) dx.$$

Nyní integrand $f(x, y(x))$ nahradíme interpolačním polynomem $P_{k-1}(x)$ stupně nejvýše $k - 1$, kde $k \geq 1$. Podle jeho výběru rozlišíme dva případy.

1. Polynom prochází body $[x_{i+1-k}, f_{i+1-k}], \dots, [x_{i-1}, f_{i-1}], [x_i, f_i]$. Dostaneme *explicitní* k -krokové metody, které se nazývají *Adamsovy*¹-*Bashforthovy*² *metody*. Značí se ABk . Uvedeme jejich vzorce pro $k = 1, 2, 3, 4, 5$:

$$y_{i+1} = y_i + hf_i, \quad (\text{vlastně Eulerova explicitní metoda})$$

$$y_{i+1} = y_i + \frac{h}{2}(3f_i - f_{i-1}),$$

$$y_{i+1} = y_i + \frac{h}{12}(23f_i - 16f_{i-1} + 5f_{i-2}),$$

$$y_{i+1} = y_i + \frac{h}{24}(55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}),$$

$$y_{i+1} = y_i + \frac{h}{720}(1\,901f_i - 2\,774f_{i-1} + 2\,616f_{i-2} - 1\,274f_{i-3} + 251f_{i-4}).$$

Řád p je stejný jako počet kroků k , tj. $p = k$.

¹**John Couch Adams** (1819–1892) (čti edems) — britský matematik a astronom. Roku 1845 předpověděl na základě matematických výpočtů (nezávisle na francouzském matematiku **Urbainu Jeanu Josephu Le Verrierovi**) existenci a polohu planety Neptun.

²**Francis Bashforth** (1819–1912) (čti bešfors) — anglický matematik a balistik.

2. Polynom prochází body $[x_{i+2-k}, f_{i+2-k}], \dots, [x_i, f_i], [x_{i+1}, f_{i+1}]$. Dostaneme *implicitní* l -krokové metody, kde $l = \max(1, k - 1)$, které se nazývají *Adamsovy-Moultonovy*¹ metody. Značí se AM_k . Uvedeme jejich vzorce pro $k = 1, 2, 3, 4, 5$ (první dvě metody jsou obě jednokrokové):

$$y_{i+1} = y_i + hf_{i+1}, \quad (\text{vlastně Eulerova implicitní metoda})$$

$$y_{i+1} = y_i + \frac{h}{2}(f_{i+1} + f_i), \quad (\text{známá jako lichoběžníkové pravidlo})$$

$$y_{i+1} = y_i + \frac{h}{12}(5f_{i+1} + 8f_i - f_{i-1}),$$

$$y_{i+1} = y_i + \frac{h}{24}(9f_{i+1} + 19f_i - 5f_{i-1} + f_{i-2}),$$

$$y_{i+1} = y_i + \frac{h}{720}(251f_{i+1} + 646f_i - 264f_{i-1} + 106f_{i-2} - 19f_{i-3}).$$

Pro řád p platí $p = k$, tj. kromě první metody je o jedničku větší než počet kroků l .

Metody AB_k i AM_k jsou D-stabilní a konzistentní, tedy konvergentní. Pouze metody AM_1 a AM_2 jsou A-stabilní. Pro dané k je velikost oblasti absolutní stability metody AM_k značně větší než u metody AB_k . S rostoucím k se oblasti absolutní stability zmenšují.

¹**Forest Ray Moulton** (1872–1952) — americký astronom, zabýval se aplikacemi matematiky v astronomii.

Metody prediktor-korektor

Tyto metody používají dvojici lineárních více krokových metod. Jednu explicitní Adamsovu-Bashforthovu metodu (*prediktor*) a jednu implicitní Adamsovu-Moultonovu metodu (*korektor*). Prediktorem (explicitní metodou) se předpoví první přiblížení $y_{i+1}^{(0)}$ hodnoty y_{i+1} , korektorem (implicitní metodou) se iterováním vylepší, tj. získá se $y_{i+1}^{(1)}$, $y_{i+1}^{(2)}$ atd. Většinou se provádí jen jedna iterace. Iterování se provádí tak, že se předpovězená hodnota dosadí do pravé strany korektoru (implicitní metoda) a výsledek, tj. levá strana je novou hodnotou pro y_{i+1} .

Podrobnějším rozbořem lokální diskretizační chyby lze zjistit, že je-li řád prediktoru q a řád korektoru p , pak řád této dvojice je $\min(p, q + 1)$, viz [9, str. 49]. Proto se nejčastěji volí $q = p$ nebo $q + 1 = p$. V obou případech je totiž řád dvojice p , tj. stejný jako řád korektoru. Oblast absolutní stability metody prediktor-korektor je větší než oblast absolutní stability prediktoru ale menší než oblast absolutní stability korektoru, ke které se přibližuje s rostoucím počtem iterací. Obecně oblasti absolutní stability dvojic AB k -AM k nejsou příliš velké a s rostoucím řádem k se zmenšují.

V dalším označíme $f_{i+1}^{(r)} = f(x_{i+1}, y_{i+1}^{(r)})$. V praxi se často používají následující dvojice:

1. Dvoukroková metoda:

Prediktor:
$$y_{i+1}^{(0)} = y_i + \frac{h}{2}(3f_i - f_{i-1}),$$
 (Adamsova-Bashforthova metoda AB2 řádu dva)

Korektor:
$$y_{i+1}^{(r+1)} = y_i + \frac{h}{2}(f_{i+1}^{(r)} + f_i), \quad r = 0, 1, 2, \dots$$
 (Adamsova-Moultonova metoda AM2 řádu dva)

2. Čtyřkroková metoda:

Prediktor:
$$y_{i+1}^{(0)} = y_i + \frac{h}{24}(55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}),$$
 (Adamsova-Bashforthova metoda AB4 řádu čtyři)

Korektor:
$$y_{i+1}^{(r+1)} = y_i + \frac{h}{24}(9f_{i+1}^{(r)} + 19f_i - 5f_{i-1} + f_{i-2}),$$

$$r = 0, 1, 2, \dots$$
 (Adamsova-Moultonova metoda AM4 řádu čtyři)

Metody prediktor-korektor se v současnosti obvykle nepoužívají s konstantním krokem a pevným řádem, ale délka kroku a případná změna řádu se řídí na základě odhadu velikosti

lokální chyby. Rovněž Adamsovy-Bashforthovy a Adamsovy-Moultonovy metody se používají s proměnným krokem.

Metody zpětného derivování

Základní myšlenkou odvození těchto metod je, že se v rovnici

$$y'(x_{i+1}) = f(x_{i+1}, y(x_{i+1}))$$

nahradí derivace $y'(x_{i+1})$ derivací interpolačního polynomu $P_k(x)$ stupně nejvýše k , kde $k \geq 1$, který je určen body $[x_{i+1-k}, y_{i+1-k}], \dots, [x_i, y_i], [x_{i+1}, y_{i+1}]$. Dostaneme *implicitní* k -krokové metody, které se nazývají *metody zpětného derivování*. Značí se BDF k (podle anglického backward differentiation formula). Uvedeme si jejich vzorce pro $k = 1, 2, 3, 4, 5, 6$.

$$y_{i+1} = y_i + hf_{i+1}, \quad (\text{vlastně Eulerova implicitní metoda})$$

$$y_{i+1} = \frac{1}{3}(4y_i - y_{i-1} + 2hf_{i+1}),$$

$$y_{i+1} = \frac{1}{11}(18y_i - 9y_{i-1} + 2y_{i-2} + 6hf_{i+1}),$$

$$y_{i+1} = \frac{1}{25}(48y_i - 36y_{i-1} + 16y_{i-2} - 3y_{i-3} + 12hf_{i+1}),$$

$$y_{i+1} = \frac{1}{137}(300y_i - 300y_{i-1} + 200y_{i-2} - 75y_{i-3} + 12y_{i-4} + 60hf_{i+1}),$$

$$y_{i+1} = \frac{1}{147}(360y_i - 450y_{i-1} + 400y_{i-2} - 225y_{i-3} + \\ + 72y_{i-4} - 10y_{i-5} + 60hf_{i+1}).$$

Řád p je stejný jako počet kroků k , tj. $p = k$.

Metody BDF k jsou D-stabilní pouze pro $k \leq 6$, tedy pro $k \geq 7$ jsou nepoužitelné. Jsou konzistentní, a tedy pro $k \leq 6$ rovněž konvergentní. Metody BDF1 a BDF2 jsou A-stabilní. Pro $3 \leq k \leq 6$ sice A-stabilní nejsou, ale oblast absolutní stability je neohraňovaná a interval absolutní stability je $(-\infty, 0)$. Jsou sice méně přesné než AM k metody, ale díky velkým oblastem absolutní stability jsou vhodné pro řešení tzv. tuhých problémů (viz oddíl 6.1.9).

6.1.9 Tuhé problémy

U některých diferenciálních rovnic se vyskytuje jev, kterému se říká *tuhost* (anglicky stiffness). Takové rovnice se nazývají *tuhé rovnice* neboli *stiff rovnice* nebo také *rovnice se silným tlumením*. Tento jev značně komplikuje jejich numerické řešení.

Pojem tuhost nemá žádnou jednoduchou a přímočarou rigorózní matematickou definici, dá se popsat mnohými různými způsoby. Typickým rysem je chování tuhých problémů, jestliže je řešíme pomocí explicitních metod Rungeho-Kutty nebo Adamsových-Bashforthových metod s pevným krokem. Dostáváme řešení, která často oscilují a obvykle neomezeně narůstají. Abychom obdrželi pomocí těchto metod přijatelná řešení, je nutné volit extrémně malý krok, což významně prodlužuje dobu řešení. Toto zvláštní chování je na první pohled nevysvětlitelné. Přitom je lze demonstrovat na velice jednoduchých rovnicích, které je možné snadno explicitně vyřešit (viz příklad 6.4).

Je možné uvést i formálnější popis tuhosti (např. u soustav lineárních diferenciálních rovnic s konstantními koeficienty souvislost s vlastními čísly), těmito otázkami se ale nebudeme zabývat. Viz např. [9, str. 56], [11, str. 26], [52, str. 525] a další.

Často má stiff rovnice jedno „význačně pomalu se měnící“ řešení, k němuž se ostatní

řešení velmi rychle přibližují pro $x \rightarrow +\infty$. Příkladem je počáteční úloha

$$y' = -100y + 100, \quad y(0) = y_0.$$

Rovnici snadno vyřešíme separací proměnných:

$$\begin{aligned} \frac{dy}{dx} &= -100(y - 1), & \ln |y - 1| &= -100x + \ln c, \\ \frac{dy}{y - 1} &= -100 dx, & y - 1 &= c e^{-100x} \\ \int \frac{dy}{y - 1} &= - \int 100 dx, & y(x) &= 1 + (y_0 - 1) e^{-100x}. \end{aligned}$$

Konstantu c jsme určili z počáteční podmínky $y(0) = y_0 = 1 + c$.

Z výsledku je vidět, že rovnice má konstantní řešení $y(x) = 1$ pro $y_0 = 1$. Ostatní řešení (pro $y_0 \neq 1$) jsou exponenciály, které se ke konstantnímu řešení rychle přibližují, protože

$$\lim_{x \rightarrow +\infty} (1 + (y_0 - 1) e^{-100x}) = 1.$$

Je třeba uvědomit si, že funkce e^{-100x} s rostoucím x velmi prudce klesá k nule.

Příklad 6.6 Najděte vzorce pro numerické řešení počáteční úlohy $y' = -100y + 100$, $y(0) = y_0$. Použijte Eulerovu explicitní a implicitní metodu. Zjistěte, jak se tato řešení chovají v závislosti na kroku h .

Řešení. Řešení budeme hledat na intervalu $\langle 0, +\infty \rangle$. Pro $f(x, y) = -100y + 100$ Eulerovou explicitní metodou postupně dostaneme užitím vzorce (6.4):

$$\begin{aligned} y_1 &= y_0 + h(-100y_0 + 100) = y_0 - 1 + 1 - 100h(y_0 - 1) = \\ &= (y_0 - 1)(1 - 100h) + 1, \end{aligned}$$

$$\begin{aligned} y_2 &= y_1 + h(-100y_1 + 100) = (y_1 - 1)(1 - 100h) + 1 = \\ &= (y_0 - 1)(1 - 100h)(1 - 100h) + 1 = (y_0 - 1)(1 - 100h)^2 + 1, \end{aligned}$$

$$\vdots$$

$$\begin{aligned} y_n &= y_{n-1} + h(-100y_{n-1} + 100) = (y_{n-1} - 1)(1 - 100h) + 1 = \\ &= (y_0 - 1)(1 - 100h)^{n-1}(1 - 100h) + 1 = (y_0 - 1)(1 - 100h)^n + 1. \end{aligned}$$

Výraz $(1 - 100h)^n$ představuje člen geometrické posloupnosti s kvocientem $q = 1 - 100h$. Pro $h > 0,02$ je $q < -1$ a posloupnost $\{q^n\}$ osciluje pro $n \rightarrow +\infty$ (posloupnost členů se

sudým indexem konverguje k $+\infty$, zatímco posloupnost členů s lichým indexem konverguje k $-\infty$). Tedy $\lim_{n \rightarrow \infty} y_n$ neexistuje (posloupnost členů se sudým indexem konverguje k $+\infty$, zatímco posloupnost členů s lichým indexem konverguje k $-\infty$ nebo naopak v závislosti na znaménku $y_0 - 1$).

Pro $h < 0,02$ bude $-1 < q < 1$, takže $\lim_{n \rightarrow \infty} (1 - 100h)^n = 0$ a $\lim_{n \rightarrow \infty} y_n = 1$. Správný výsledek, tj. takový, že $\lim_{n \rightarrow \infty} y_n = 1$, tudíž dostaneme jen pro dostatečně malý krok h , jinak je numerické řešení zcela chybné.

Eulerovou implicitní metodou užitím vzorce (6.6) dostaneme pro y_1 :

$$\begin{aligned} y_1 &= y_0 + h(-100y_1 + 100), \\ y_1(1 + 100h) &= 1 + 100h + y_0 - 1, \\ y_1 &= 1 + \frac{y_0 - 1}{1 + 100h}. \end{aligned}$$

Obdobně se odvodí, že

$$\begin{aligned}
 y_2 &= 1 + \frac{y_1 - 1}{1 + 100h} = 1 + \frac{y_0 - 1}{1 + 100h} \frac{1}{1 + 100h} = 1 + \frac{y_0 - 1}{(1 + 100h)^2}, \\
 &\vdots \\
 y_n &= 1 + \frac{y_{n-1} - 1}{1 + 100h} = 1 + \frac{y_0 - 1}{(1 + 100h)^{n-1}} \frac{1}{1 + 100h} = 1 + \frac{y_0 - 1}{(1 + 100h)^n}.
 \end{aligned}$$

Výraz $1/(1 + 100h)^n$ představuje člen geometrické posloupnosti s kvocientem $q = 1/(1 + 100h)$. Pro každé $h > 0$ je $0 < q < 1$, tedy $\lim_{n \rightarrow \infty} q^n = 0$ a $\lim_{n \rightarrow \infty} y_n = 1$. Tentokrát tudíž správný výsledek dostaneme pro libovolný krok h . ▲

Poznatek z předchozího příkladu platí obecně. Některé metody dávají pro tuhé problémy dobrý výsledek jen při velmi malém kroku h , což je z hlediska objemu výpočtů neúnosné. Říká se jim *netuhé metody* nebo *non-stiff metody*. Je pro ně typické, že mají malou oblast absolutní stability.

Tuhé rovnice se vyskytují v chemii, kde popisují průběh chemických reakcí, v elektronice, kde např. tzv. van der Polova rovnice (viz příklad 6.8) popisuje kmity oscilátoru, v kinematice a mnoha dalších důležitých aplikacích. Pro jejich numerické řešení byly

vyvinuty speciální metody, umožňující prodloužit krok, kterým se říká *tuhé metody* nebo *stiff metody*. Ty naopak nejsou vhodné pro řešení problémů, které nejsou tuhé. Tuhé metody jsou implicitní. Je důležité, aby měly neohrazenou oblast absolutní stability, což umožňuje zvětšit krok.

Všechny jednokrokové metody, které jsme uvedli, jsou až na Eulerovu implicitní metodu a AM2 metodu netuhé. Mezi tuhé metody patří semiimplicitní nebo implicitní metody typu Rungeho-Kutty. Z uvedených víckrokových metod jsou tuhé metody zpětného diferencování.

V programu Maple lze najít řadu účinných tuhých metod, např. Rosenbrockovu (jistě zobecnění metod typu Rungeho-Kutty, viz [7, str. 120], [78]) nebo skupinu metod lsode (Livermore Stiff ODE solver; založeno na kombinaci AM a BDF metod, viz [73]).

Problematika numerického řešení obyčejných diferenciálních rovnic prodělala v posledních čtyřiceti letech bouřlivý rozvoj, který byl především způsoben nástupem počítačů a jejich neustále se zvyšujícími schopnostmi. Bylo možné využívat čím dál složitější metody vyžadující velký objem výpočtů. To vyvolalo zpětně tlak na hlubší zkoumání teoretických vlastností nových metod. Nové poznatky pak umožnily dokonalejší implementaci těchto metod. Vzhledem k významu řešení diferenciálních rovnic tento vývoj rozhodně není ukončen.

6.2 Systémy obyčejných diferenciálních rovnic

Doposud jsme se zabývali numerickým řešením jedné diferenciální rovnice prvního řádu. Mnoho matematických modelů je ale tvořeno soustavou více diferenciálních rovnic. Nechť $n \in \mathbb{N}$. Systém n diferenciálních rovnic prvního řádu pro neznámé funkce $y_1(x)$, $y_2(x)$ až $y_n(x)$ s počátečními podmínkami $y_1(x_0) = y_{1,0}$, $y_2(x_0) = y_{2,0}$ až $y_n(x_0) = y_{n,0}$, kde x_0 , $y_{1,0}$, $y_{2,0}$ až $y_{n,0}$ jsou daná čísla, lze formálně zapsat takto:

$$\begin{aligned}
 y_1' &= f_1(x, y_1, \dots, y_n), & y_1(x_0) &= y_{1,0}, \\
 y_2' &= f_2(x, y_1, \dots, y_n), & y_2(x_0) &= y_{2,0}, \\
 &\vdots & & \\
 y_n' &= f_n(x, y_1, \dots, y_n), & y_n(x_0) &= y_{n,0}.
 \end{aligned}
 \tag{6.14}$$

Přitom funkce $f_1(x, y_1, \dots, y_n)$ až $f_n(x, y_1, \dots, y_n)$ popisují pravé strany diferenciálních rovnic.

Předchozí zápis je komplikovaný. Proto systém zapíšeme pomocí tzv. *vektorového*

neboli *maticového zápisu*. Označíme:

$$\mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \\ \vdots \\ y_n(x) \end{pmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{pmatrix} f_1(x, y_1, \dots, y_n) \\ f_2(x, y_1, \dots, y_n) \\ \vdots \\ f_n(x, y_1, \dots, y_n) \end{pmatrix}, \quad \mathbf{y}_0 = \begin{pmatrix} y_{1,0} \\ y_{2,0} \\ \vdots \\ y_{n,0} \end{pmatrix}.$$

Definujme ještě

$$\mathbf{y}'(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \\ \vdots \\ y_n(x) \end{pmatrix}' = \begin{pmatrix} y_1'(x) \\ y_2'(x) \\ \vdots \\ y_n'(x) \end{pmatrix}.$$

Cauchyovu počáteční úlohu pro systém (6.14) lze pak stručně zapsat takto:

$$\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0. \quad (6.15)$$

Použití vektorového zápisu si ukážeme na příkladu.

Příklad 6.7 Napište ve vektorovém zápisu soustavu tří diferenciálních rovnic

$$y_1' = -10y_1 + 10y_2, \quad y_2' = -y_1y_3 + 28y_1 - y_2, \quad y_3' = y_1y_2 - \frac{8}{3}y_3$$

s počátečními podmínkami

$$y_1(0) = -8, \quad y_2(0) = 8, \quad y_3(0) = 27.$$

Řešení. Protože $n = 3$, označíme:

$$\mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \\ y_3(x) \end{pmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{pmatrix} -10y_1 + 10y_2 \\ -y_1y_3 + 28y_1 - y_2 \\ y_1y_2 - \frac{8}{3}y_3 \end{pmatrix}, \quad \mathbf{y}_0 = \begin{pmatrix} -8 \\ 8 \\ 27 \end{pmatrix}.$$

Pak platí

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} -10y_1 + 10y_2 \\ -y_1y_3 + 28y_1 - y_2 \\ y_1y_2 - \frac{8}{3}y_3 \end{pmatrix}, \quad \begin{pmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \end{pmatrix} = \begin{pmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \end{pmatrix} = \begin{pmatrix} -8 \\ 8 \\ 27 \end{pmatrix},$$

čili stručně

$$\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}_0.$$

Tento systém se nazývá *Lorenzův*¹. Vyskytuje se v teoretické meteorologii. Explicitně ho řešit nelze a při numerickém řešení je třeba postupovat opatrně. Řešení lze znázornit jako prostorovou křivku o parametrických rovnicích $\mathbf{y} = \mathbf{y}(x)$ (souřadnicové osy jsou označené y_1 , y_2 a y_3 ,). Přestože je tento systém poměrně jednoduchý, chování jeho řešení je velmi složité, tzv. chaotické, jedno z řešení, tzv. *podivný atraktor*, má fraktální strukturu, viz animace na obr. 6.16 (křivka mění barvu, aby bylo lépe vidět, jak komplikovaný je její průběh).

Podobné vlastnosti mají diferenciální rovnice popisující např. dynamické systémy, které jsou modely zemské atmosféry. Proto je předpovídání počasí tak obtížná záležitost a dlouhodobější předpovědi jsou často velmi nespolehlivé. ▲

¹**Edward Norton Lorenz** (1917–2008) — americký matematik a meteorolog, jeden z průkopníků teorie chaosu.

Obr. 6.16: Lorenzův podivný atraktor

O vlastnostech řešení systémů diferenciálních rovnic lze dokázat obdobné výsledky jako o jedné diferenciální rovnici. Rovněž numerické metody a pojmy s nimi související se téměř beze zbytku přenášejí na systémy. Všechny vzorce z předchozích kapitol pro jednokrokové i víceurokové metody zůstávají v platnosti, pokud použijeme vektorový zápis.

Označme

$$\mathbf{y}_i = \begin{pmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{n,i} \end{pmatrix}, \quad i = 0, 1, \dots, m.$$

Pak např. explicitní Eulerova metoda má tvar

$$\mathbf{y}_{i+1} = \mathbf{y}_i + h \mathbf{f}(x_i, \mathbf{y}_i), \quad i = 0, 1, \dots, m-1.$$

Složky vektorů \mathbf{y}_i dávají pro každou neznámou $y_1(x)$ až $y_n(x)$ hodnoty v uzlových bodech x_0, x_1, \dots, x_m .

6.3 Diferenciální rovnice vyšších řádů

Kromě diferenciálních rovnic prvního řádu jsou v aplikacích velmi důležité i diferenciální rovnice vyšších řádů. Ukážeme si, jak lze jednu takovou rovnici vyššího řádu přepsat na soustavu více rovnic prvního řádu. Podobně lze postupovat i u systémů diferenciálních rovnic vyšších řádů, pokud jsou v explicitním tvaru (rovnice jsou vyřešené vzhledem k nejvyšší derivaci).

Nechť $n \in \mathbb{N}$. Cauchyova počáteční úloha pro diferenciální rovnici n -tého řádu s neznámou $y(x)$ má obecný tvar

$$\begin{aligned} y^{(n)} &= f(x, y, y', \dots, y^{(n-1)}), \\ y(x_0) &= y_{1,0}, \quad y'(x_0) = y_{2,0}, \dots, y^{(n-1)}(x_0) = y_{n,0}, \end{aligned} \tag{6.16}$$

kde $x_0, y_{1,0}, y_{2,0}, \dots, y_{n,0}$ jsou daná čísla a funkce f popisuje pravou stranu rovnice.

Označme

$$y_1(x) = y(x), \quad y_2(x) = y'(x), \dots, y_n(x) = y^{(n-1)}(x).$$

Pak platí

$$\begin{aligned}
 y_1' &= y_2, & y_1(x_0) &= y_{1,0}, \\
 y_2' &= y_3, & y_2(x_0) &= y_{2,0}, \\
 &\vdots & & \\
 y_{n-1}' &= y_n, & y_{n-1}(x_0) &= y_{n-1,0}, \\
 y_n' &= f(x, y_1, \dots, y_n), & y_n(x_0) &= y_{n,0}.
 \end{aligned}
 \tag{6.17}$$

Jestliže je $y(x)$ řešením rovnice (6.16), snadno se ověří, že potom sloupcový vektor $(y(x), y'(x), \dots, y^{(n-1)}(x))^T$ je řešením soustavy (6.17).

Naopak, je-li vektor $(y_1(x), y_2(x), \dots, y_n(x))^T$ řešením soustavy (6.17), je $y_1(x)$ řešením rovnice (6.16).

Označíme-li tedy

$$f(x, \mathbf{y}) = \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_n \\ f(x, y_1, \dots, y_n) \end{pmatrix},$$

lze rovnici (6.16) nahradit systémem (6.17), jehož vektorový zápis je

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0.$$

Ten lze numericky řešit některou z dříve uvedených metod. Poznamenejme, že existují i speciální numerické metody určené pro rovnice vyšších řádů.

Příklad 6.8 Přepište na systém ve vektorovém zápisu diferenciální rovnici druhého řádu

$$y'' + \mu(y^2 - 1)y' + y = 0, \quad y(0) = 1, \quad y'(0) = 0,$$

kde $\mu > 0$ je reálný parametr.

Řešení. Označíme

$$y_1 = y, \quad y_2 = y'.$$

Pak platí

$$\begin{aligned} y_1' &= y_2, & y_{1,0} &= 1, \\ y_2' &= -y_1 + \mu(1 - y_1^2)y_2, & y_{2,0} &= 0. \end{aligned}$$

Označme

$$\mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \end{pmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{pmatrix} y_2 \\ -y_1 + \mu(1 - y_1^2)y_2 \end{pmatrix}, \quad \mathbf{y}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Pak platí

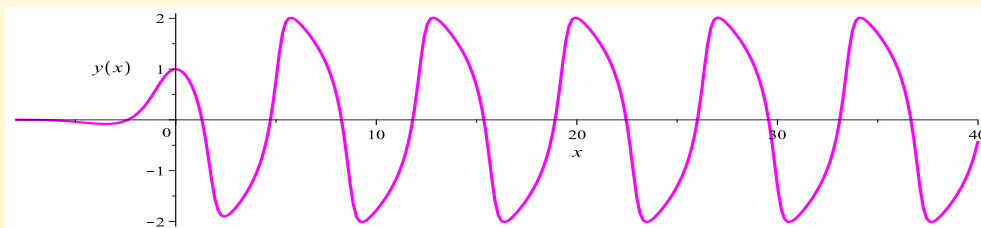
$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} y_2 \\ -y_1 + \mu(1 - y_1^2)y_2 \end{pmatrix}, \quad \begin{pmatrix} y_1(0) \\ y_2(0) \end{pmatrix} = \begin{pmatrix} y_{1,0} \\ y_{2,0} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

čili stručně

$$\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}_0. \quad \blacktriangle$$

Tato rovnice se nazývá *van der Polova*¹. V dynamice tato rovnice popisuje nekonzervativní oscilátor s nelineárním tlumením. Má použití v elektrotechnice a biologii. Při jejím studiu došlo k jednomu z prvních případů zaznamenání *deterministického chaosu*. Řešení na některých částech intervalu rychle oscilují, takže pro numerické výpočty jsou vhodné stiff metody.

¹**Balthasar van der Pol** (1889–1959) — holandský fyzik a elektroinženýr, pracoval pro Philips.



Obr. 6.17: Řešení van der Polovy rovnice

Na obr. 6.17 je znázorněno numericky získané řešení uvedeného počátečního problému pro $\mu = 1,5$. Van der Polova rovnice je tzv. *autonomní*, což znamená, že v rovnici se explicitně nevyskytuje nezávisle proměnná x . Řešení takových rovnic se obvykle znázorňují ve *fázové rovině*, jejíž souřadnicové osy jsou y_1 a y_2 , tedy vlastně y a y' . Řešení jsou znázorněna jako *trajektorie*, což jsou křivky s parametrickými rovnicemi $[y_1(t), y_2(t)]$ neboli $[y(t), y'(t)]$. Uzavřeným trajektoriím odpovídají periodická řešení. Lze dokázat, že van der Polova rovnice má jedinou uzavřenou trajektorii a všechny ostatní trajektorie se na ni s rostoucím x zvenku nebo zevnitř navíjejí.

Animace na obr. 6.18 s částmi sedmi trajektorií znázorňuje toto chování. Je vidět, jak body trajektorií probíhají některé úseky velmi rychle a jiné zase pomalu a rychle se přibližují k uzavřené trajektorii. Počátek $[0, 0]$ je jednobodová trajektorie odpovídající řešení $y_1(x) \equiv 0$, $y_2(x) \equiv 0$. Protože tuto rovnice nelze explicitně řešit, byly obrázek i animace získány numericky pro $\mu = 1,5$.

Složitě chování Lorenzova podivného atraktoru i řešení van der Polovy rovnice ukazují, jak účinné jsou soudobé numerické metody pro řešení obyčejných diferenciálních rovnic.

Obr. 6.18: Trajektorie van der Polovy rovnice

Příklad 6.9 Přepište na systém ve vektorovém zápisu diferenciální rovnici třetího řádu

$$y''' = xy + (y')^2 - yy'', \quad y(1) = 2, \quad y'(1) = -1, \quad y''(1) = 0.$$

Řešení. Označíme

$$y_1 = y, \quad y_2 = y', \quad y_3 = y''.$$

Pak platí

$$\begin{aligned} y_1' &= y_2, & y_{1,0} &= 2, \\ y_2' &= y_3, & y_{2,0} &= -1, \\ y_3' &= xy_1 + y_2^2 - y_1y_3, & y_{3,0} &= 0. \end{aligned}$$

Označme

$$\mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \\ y_3(x) \end{pmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{pmatrix} y_2 \\ y_3 \\ xy_1 + y_2^2 - y_1y_3 \end{pmatrix}, \quad \mathbf{y}_0 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}.$$

Pak platí

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} y_2 \\ y_3 \\ xy_1 + y_2^2 - y_1y_3 \end{pmatrix}, \quad \begin{pmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \end{pmatrix} = \begin{pmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix},$$

čili stručně

$$\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}_0. \quad \blacktriangle$$

V této kapitole jsme se zabývali numerickým řešením Cauchyovy počáteční úlohy, kdy z nekonečně mnoha řešení dané diferenciální rovnice nebo systému se vybere jedno řešení pomocí počátečních podmínek. V praxi se ale setkáváme i s jinými typy podmínek, které určují jediné řešení. Z nich nejdůležitější jsou *okrajové podmínky*. Pro numerické řešení těchto tzv. *okrajových úloh* existují speciální metody, např. metoda střelby, diferenční metoda, metoda konečných objemů a metoda konečných prvků. Viz např. [9, 11, 52, 57].

Pojmy k zapamatování

- obyčejné diferenciální rovnice prvního řádu v implicitním a explicitním tvaru
- řešení diferenciální rovnice
- počáteční podmínka, Cauchyova počáteční úloha
- existence a jednoznačnost řešení počáteční úlohy
- princip numerického řešení počáteční úlohy
- jednokrokové a vícekové metody
- tvar obecné explicitní jednokrokové metody
- tvar obecné implicitní jednokrokové metody
- explicitní a implicitní Eulerova metoda
- lokální diskretizační, lokální a globální chyba a řád jednokrokové metody
- konvergence jednokrokové metody
- stabilita jednokrokové metody, oblast a interval absolutní stability
- metody Rungeho-Kutty a jejich princip, stupeň a řád, Butcherova tabulka

- tvar obecné lineární více krokové metody
- explicitní a implicitní více krokové metody
- první a druhý charakteristický polynom více krokové metody
- lokální diskretizační a globální chyba a řád více krokové metody
- konzistentnost, D-stabilita a konvergence více krokových metod
- stabilita více krokové metody, oblast a interval absolutní stability
- Adamsovy-Bashforthovy a Adamsovy-Moultonovy metody a jejich princip
- metody prediktor-korektor a jejich princip
- metody zpětného derivování a jejich princip
- tuhé (stiff) problémy
- systémy obyčejných diferenciálních rovnic prvního řádu, Cauchyova počáteční úloha
- metody numerického řešení systémů diferenciálních rovnic
- obyčejné diferenciální rovnice vyšších řádů, Cauchyova počáteční úloha
- přepis diferenciální rovnice vyššího řádu na systém prvního řádu

Kontrolní otázky

1. Vysvětlete pojmy a ilustруйте je na příkladech: ODR 1. řádu v implicitním tvaru, v explicitním tvaru, řešení ODR 1. řádu, směrové pole ODR.
2. Vysvětlete pojem Cauchyova počáteční úloha a uveďte podmínky, za kterých má tato úloha jediné řešení.
3. Co rozumíme numerickým řešením počáteční úlohy?
4. Vysvětlete princip jednokrokových a více krokových metod.
5. Jaké problémy mohou nastat při použití více krokové metody?
6. Napište vzorce pro implicitní a explicitní Eulerovu metodu a aplikujte je na konkrétním příkladu.
7. Napište vzorec obecné explicitní a implicitní jednokrokové metody.
8. Vysvětlete rozdíl v použití explicitních a implicitních jednokrokových metod.
9. Vysvětlete pojmy lokální diskretizační chyba, lokální chyba a globální chyba jednokrokové metody.
10. Co je to řád metody, jak souvisí s lokální diskretizační chybou jednokrokové metody?

11. Co znamená, že jednokroková metoda je konvergentní? Co musí být splněno, aby metoda byla konvergentní?
12. Vysvětlete, co je oblast a interval absolutní stability jednokrokové metody. Co znamená, že taková metoda je A-stabilní?
13. Na jakém principu jsou založeny metody Rungeho-Kutty?
14. Vysvětlete postup při použití explicitní metody Rungeho-Kutty 3. řádu.
15. Co je to stupeň a řád metody Rungeho-Kutty? Jaký je mezi nimi vztah (u explicitních metod)?
16. Napište tvar obecné lineární k -krokové metody. Vysvětlete, kdy půjde o předpis pro explicitní a kdy pro implicitní metodu.
17. Vysvětlete pojmy lokální diskretizační chyba, globální chyba a řád vícekrokové metody.
18. Co znamená, že vícekroková metoda je konvergentní? Co musí být splněno, aby metoda byla konvergentní?
19. Co je oblast absolutní stability lineární vícekrokové metody? Co znamená, že taková metoda je A-stabilní?
20. Uveďte příklady nejvýznamnějších lineárních vícekrokových metod.
21. Na jakém principu jsou založeny metody prediktor-korektor?

22. Co jsou to tuhé neboli stiff problémy?
23. Zapište, jak vypadá obecný systém n diferenciálních rovnic prvního řádu s počátečními podmínkami, a popište, jak se numericky řeší.
24. Uveďte postup při numerickém řešení diferenciální rovnice vyššího řádu.
25. Jaký je rozdíl mezi numerickým řešením jedné diferenciální rovnice 1. řádu a soustavy diferenciálních rovnic 1. řádu?

Testy ke kapitole 6

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Která z následujících diferenciálních rovnic 1. řádu je v explicitním tvaru?

$$\ln(x^2 - xy + y') = y'$$

$$y' = xy - \cos(xy')$$

$$y' = -\frac{\sqrt{1-y}}{y\sqrt{1-x^2}}$$

$$xy' - y = \sqrt{x^2 - xy'}$$

2. (1b.) Platí obecně, že čím menší budou kroky h_i , tím přesnější bude numerické řešení diferenciální rovnice?

Ne.

Ano.

3. (1b.) Vyberte definiční obor pravé strany diferenciální rovnice $y' = \frac{3}{\sqrt{4x - 5y}}$.

$$D(f) = \left\{ [x, y] \in \mathbb{R}^2 : x \neq \frac{5}{4} \right\}.$$

$$D(f) = \left\{ [x, y] \in \mathbb{R}^2 : x = \frac{5}{4} \right\}.$$

$$D(f) = \left\{ [x, y] \in \mathbb{R}^2 : y \leq \frac{4}{5}x \right\}.$$

$$D(f) = \left\{ [x, y] \in \mathbb{R}^2 : y < \frac{4}{5}x \right\}.$$

4. (1b.) Vyberte vzorec explicitní Eulerovy metody s proměnným krokem h_i pro počáteční úlohu $y' = x - y^2$, $y(x_0) = y_0$.

$$y_{i+1} = y_i + h_{i+1}(x_{i+1} - y_{i+1}^2).$$

$$y_{i+1} = y_i + h_{i+1}(x_i - y_i^2).$$

$$y_{i+1} = y_i + h_i(x_i - y_i^2).$$

$$y_{i+1} = y_i + h_i(x_{i+1} - y_{i+1}^2).$$

5. (1b.) Chyba, která vzniká kumulací skutečných chyb z jednotlivých kroků výpočtu, se nazývá

globální chyba.

lokální chyba.

lokální diskretizační
chyba.

6. (1b.) Metody Rungeho-Kutty patří do skupiny

jednokrokových metod.

vícekrokových metod.

7. (1b.) Vyberte předpis pro obecnou implicitní lineární k -krokovou metodu s konstantním krokem.

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}, y_{i+1}; f).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}, y_{i+1}; f).$$

$$y_{i+1} = -a_1 y_i - \cdots - a_k y_{i+1-k} + h(b_0 f_{i+1} + \cdots + b_k f_{i+1-k}), b_0 \neq 0.$$

$$y_{i+1} = -a_1 y_i - \cdots - a_k y_{i+1-k} + h(b_1 f_i + \cdots + b_k f_{i+1-k}).$$

8. (1b.) Lineární vícekroková metoda se nazývá konzistentní, jestliže

je konvergentní.

je implicitní.

je explicitní.

je řádu alespoň jedna.

9. (1b.) Mezi víceřkové metody patří metody prediktor-korektor. Na jakém principu jsou založeny?

Na integraci rovnice $y'(x) = f(x, y(x))$ s počáteční podmínkou $y(x_0) = y_0$ na intervalu $\langle x_i, x_{i+1} \rangle$ a následné náhradě integrandu $f(x, y(x))$ vhodným interpolačním polynomem.

Tyto metody používají dvojici lineárních víceřkových metod. Jednou metodou se předpoví hodnota y_{i+1} a druhou metodou se získaná hodnota vylepší.

V rovnici $y'(x_{i+1}) = f(x_{i+1}, y(x_{i+1}))$ se nahradí derivace $y'(x_{i+1})$ derivací interpolačního polynomu stupně nejvýše k , určeného vhodnými body.

10. (1b.) Co platí o systémech obyčejných diferenciálních rovnic?

Pro systémy diferenciálních rovnic neexistuje obdoba víceřkových numerických metod.

Pro systémy diferenciálních rovnic lze odvodit obdobné numerické metody jako pro jednu diferenciální rovnici.

Vícekové numerické metody musí používat aspoň tolik kroků, kolik je počet rovnic v systému.

Systémy diferenciálních rovnic nelze řešit jednokrokovými metodami.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (1b.) Vyberte, kterou z uvedených počátečních podmínek žádné řešení diferenciální

rovnice $y' = \sqrt{y^2 - \frac{x}{2}}$ nespĺňuje.

$$y(1) = 3.$$

$$y(2) = 1$$

$$y(3) = 1.$$

$$y(1) = 2.$$

2. (1b.) Vyberte vzorec implicitní Eulerovy metody s proměnným krokem h_i pro počáteční úlohu $y' = 2x^2 + x - y$, $y(x_0) = y_0$.

$$y_{i+1} = y_i + h_{i+1}(2x_i^2 + x_i - y_i).$$

$$y_{i+1} = y_i + h_i(2x_{i+1}^2 + x_{i+1} - y_{i+1}).$$

$$y_{i+1} = y_i + h_{i+1}(2x_i^2 + x_i - y_{i+1}).$$

$$y_{i+1} = y_i + h_{i+1}(2x_{i+1}^2 + x_{i+1} - y_{i+1}).$$

3. (1b.) Numerická metoda pro řešení obyčejných diferenciálních rovnic, která k výpočtu hodnoty y_{i+1} používá hodnoty y_i a y_{i-1} , se nazývá
- | | |
|---------------------|----------------------|
| tříkroková metoda. | explicitní metoda. |
| dvoukroková metoda. | jednokroková metoda. |
4. (1b.) Co je nevýhodou jednokrokových implicitních metod pro řešení obyčejných diferenciálních rovnic?

Jsou méně přesné než explicitní metody. Navíc je potřeba delší čas na jejich řešení, protože neznámá hodnota y_{i+1} se vyskytuje ve vzorcích těchto metod na obou stranách rovnice.

Zpravidla konvergují k přesnému řešení pomaleji než explicitní metody.

Velikost chyby, které se dopustíme, je o řád větší, než když použijeme explicitní metodu.

Neznámá hodnota y_{i+1} se vyskytuje ve vzorcích těchto metod na obou stranách rovnice. Po dosazení za y_i tak obdržíme obecně nelineární rovnici, která se zpravidla musí řešit numericky.

5. (1b.) Souvisí spolu řád metody p a lokální diskretizační chyba metody?

Ano. Čím menší je řád p metody, tím menší je lokální diskretizační chyba.

Ano. Je-li metoda řádu jedna, pak je lokální diskretizační chyba nejmenší.

Ano. Uvažujeme-li metodu s ekvidistantní délkou kroku h , pak platí, že čím je metoda vyššího řádu p , tím menší je lokální diskretizační chyba.

Ne. Jde o nezávislé pojmy.

6. (1b.) Co je typické pro diferenciální rovnice, které se označují jako tuhé neboli stiff?

Na jejich numerické řešení je třeba použít tzv. tuhé neboli stiff metody.

Nelze je numericky řešit.

Jejich numerické řešení je bezproblémové, lze použít v podstatě libovolnou metodu.

Na jejich numerické řešení je třeba použít explicitní metody s malou oblastí absolutní stability.

7. (1b.) Obecný tvar s -stupňové metody Rungeho-Kutty, kde $s \in \mathbb{N}$, $s \geq 1$, s konstantním krokem h má tvar

$$y_{i+1} = y_i + h(\gamma_1 k_1 + \gamma_2 k_2 + \cdots + \gamma_s k_s), \quad i = 0, \dots, n-1.$$

Co víme o koeficientech $\gamma_1, \dots, \gamma_s$?

Platí $\gamma_1 + \gamma_2 + \cdots + \gamma_s = 1$, kde $\gamma_r \leq 0$ pro $r = 1, \dots, s$.

Platí $\gamma_1 + \gamma_2 + \cdots + \gamma_s = 1$, kde $\gamma_r \geq 0$ pro $r = 1, \dots, s$.

Platí $\gamma_1 + \gamma_2 + \cdots + \gamma_s = 1$.

Platí $\gamma_1 \cdot \gamma_2 \cdot \cdots \cdot \gamma_s = 1$.

O koeficientech neplatí nic, mohou to být libovolná kladná nebo záporná čísla.

8. (1b.) Co znamená symbol f_i ve vzorcích lineárních více krokových metod?

$$f_i = f(x_i, y_i).$$

$$f_i = f(x_i, y(x_i)).$$

$$f_i = f(x_{i-1}, y_{i-1}).$$

$$f_i = f(x_{i+1}, y_{i+1}).$$

9. (1b.) Co platí o numerickém řešení počáteční úlohy pro diferenciální rovnice vyšších řádů?

Pro numerické řešení počáteční úlohy pro diferenciální rovnice vyšších řádů je nutné použít zcela speciální metody.

Numerické metody pro řešení počátečních úloh pro diferenciální rovnice vyšších řádů neexistují, protože jde o velmi špatně podmíněné úlohy.

Diferenciální rovnice vyšších řádů je možné přepsat na systém diferenciálních rovnic prvního řádu a použít numerické metody pro řešení takových systémů.

Počáteční úlohy pro diferenciální rovnice vyšších řádů lze numericky řešit pouze vícekrokovými metodami, přičemž počet kroků musí odpovídat řádu těchto rovnic.

10. (1b.) Lineární vícekroková metoda je konvergentní,
právě když je explicitní a implicitní. právě když je konzistentní a D-stabilní.
jestliže je D-stabilní a explicitní. jestliže je konzistentní a implicitní.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (1b.) Rozhodněte, zda některé řešení diferenciální rovnice $y' - y \operatorname{tg} x = \frac{1}{\cos x}$ splňuje počáteční podmínku $y(0) = 0$.

Ano. Funkce $f(x, y) = y \operatorname{tg} x + \frac{1}{\cos x}$ je v bodě $[0,0]$ definovaná a spojitá.

Ne. Po dosazení počáteční podmínky dostaneme rovnici $y' = 1$, což není diferenciální rovnice.

Ne, protože $\cos 0 = 1$, a ne nula.

2. (1b.) Vyberte předpis explicitní Eulerovy metody s konstantním krokem h pro řešení diferenciální rovnice $y' = f(x, y)$.

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}).$$

$$y_{i+1} = y_i + hf(x_i^2, y_i^2).$$

$$y_{i+1} = y_0 + hf(x_i, y_i).$$

$$y_{i+1} = y_i + hf(x_i, y_i).$$

3. (1b.) Rozhodněte, zda je tvrzení pravdivé. Globální chyba metody je rovna součtu lokálních chyb.

Ano.

Ne.

4. (1b.) Vyberte předpis pro obecnou implicitní jednokrokovou metodu s proměnným krokem.

$$y_{i+1} = y_{i-1} + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}, y_{i+1}; f).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_{i+1}, h_{i+1}; f).$$

5. (1b.) V jednokrokových metodách se hodnota y_{i+1} počítá z hodnoty y_i , která odpovídá bezprostředně předcházejícímu uzlu. pomocí libovolné předcházející hodnoty $y_i, i = 0, \dots, n - 1$. výhradně pomocí hodnoty y_0 .

6. (1b.) Jaký je vztah mezi diferenciálními rovnicemi prvního řádu v implicitním tvaru a implicitními metodami pro numerické řešení diferenciálních rovnic?

Jsou to jen různé názvy pro tentýž pojem.

Jde o dva zcela různé pojmy.

Diferenciální rovnice v implicitním tvaru lze řešit pouze implicitními metodami.

Implicitní metody nelze použít pro řešení diferenciálních rovnic v implicitním tvaru.

7. (1b.) K čemu obvykle dojde při řešení tuhých diferenciálních rovnic pomocí explicitní metody Rungeho-Kutty s pevným krokem?

Konvergence je pomalejší, záleží na řádu metody, ale přesnost je obvykle dostatečná.

Konvergence je zajištěna při libovolně dlouhém kroku.

Nelze volit příliš malý krok, jinak dochází k oscilování numerického řešení.

Pokud nezvolíme velmi malý krok, numerická řešení často oscilují a jsou zcela znehodnocena.

8. (1b.) Mezi vícezkrové metody patří Adamsovy metody. Na jakém principu jsou založeny?

Na integraci rovnice $y'(x) = f(x, y(x))$ s počáteční podmínkou $y(x_0) = y_0$ na intervalu $\langle x_i, x_{i+1} \rangle$ a následně náhradě integrandu $f(x, y(x))$ vhodným interpolačním polynomem.

Tyto metody používají dvojici lineárních vícezkrových metod. Jednou metodou se předpoví hodnota y_{i+1} a druhou metodou se získaná hodnota vylepší.

V rovnici $y'(x_{i+1}) = f(x_{i+1}, y(x_{i+1}))$ se nahradí derivace $y'(x_{i+1})$ derivací interpolačního polynomu stupně nejvýše k , určeného vhodnými body.

9. (1b.) Numerická metoda pro řešení obyčejných diferenciálních rovnic má řád p , jestliže p je největší přirozené číslo, pro nějž platí (lte_i značí lokální diskretizační chybu, e_i globální diskretizační chybu):

$$e_i = O(h^{p+1}), \quad e_i = O(h^p), \quad lte_i = O(h^{p+1}), \quad lte_i = O(h^p).$$

10. (1b.) V definici lokální diskretizační chyby lte_i se předpokládá, že

hodnotu y_{i+1} počítáme z hodnot y_i a $y(x_{i+1})$, je to tzv. globalizační předpoklad.

hodnotu y_{i+1} počítáme z přesných hodnot $y(x_i)$ a $y(x_{i+1})$, je to tzv. lokalizační předpoklad.

hodnotu y_{i+1} počítáme z odhadu hodnoty y_i .

hodnotu y_{i+1} počítáme z hodnoty $y(x_0)$, která je jediná zcela přesná.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 4

1. (1b.) Jestliže řešíme diferenciální rovnici 1. řádu numericky, pak

řešením je funkce $y = f(x)$ definovaná na intervalu $\langle a, b \rangle$.

řešením je přibližná hodnota přesné hodnoty $y(b)$ v koncovém bodě intervalu $\langle a, b \rangle$, na kterém hledáme řešení.

řešením je přesná hodnota $y(b)$ v koncovém bodě intervalu $\langle a, b \rangle$, na kterém hledáme řešení.

řešením jsou přibližné hodnoty y_i přesných hodnot $y(x_i)$ v uzlových bodech x_i , $i = 0, \dots, n$, na intervalu $\langle a, b \rangle$, kde řešení hledáme.

2. (1b.) Rozhodněte, zda je tvrzení pravdivé. Lokální diskretizační chyba metody je skutečná chyba, které se dopouštíme v jednom kroku.

Ne.

Ano.

3. (1b.) Vyberte předpis implicitní Eulerovy metody s konstantním krokem h pro řešení diferenciální rovnice $y' = f(x, y)$.

$$y_{i+1} = y_i + hf(x_i, y_i).$$

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}).$$

$$y_{i+1} = y_i + hf(x_i^2, y_i^2).$$

$$y_{i+1} = y_0 + hf(x_i, y_i).$$

4. (1b.) Je-li numerická metoda pro řešení diferenciálních rovnic konvergentní, pak existuje takový index i , $i = 2, \dots, n - 1$, že přibližné hodnoty y_i splynou s přesnými hodnotami $y(x_i)$.

se při zmenšujícím konstantním kroku h přibližné hodnoty y_i přibližují k přesným hodnotám $y(x_i)$.

se při zmenšujícím konstantním kroku h přibližné hodnoty y_i od jistého indexu přibližují k nule.

5. (1b.) S tzv. Butcherovou tabulkou se můžeme setkat při řešení diferenciálních rovnic pomocí metod Rungeho-Kutty.

Butcherovou metodou.

metodou prediktor-korektor.

vícekrokovými metodami.

6. (1b.) Vyberte předpis pro obecnou explicitní jednokrokovou metodu s proměnným krokem.

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}, y_{i+1}; f).$$

$$y_{i+1} = y_{i-1} + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_{i+1}, h_{i+1}; f).$$

7. (1b.) Co je typické pro chování řešení tuhých diferenciálních rovnic?

Všechna řešení velmi rychle oscilují.

Řešení mají do nekonečna nevládní limity plus nebo minus nekonečno.

Často existuje jedno význačné pomalu se měnící řešení, k němuž se ostatní řešení rychle přibližují.

Všechna řešení se rychle blíží k nule.

8. (1b.) Mezi vícezkrokové metody patří metody zpětného derivování. Na jakém principu jsou založeny?

V rovnici $y'(x_{i+1}) = f(x_{i+1}, y(x_{i+1}))$ se nahradí derivace $y'(x_{i+1})$ derivací interpolačního polynomu stupně nejvýše k , určeného vhodnými body.

Na integraci rovnice $y'(x) = f(x, y(x))$ s počáteční podmínkou $y(x_0) = y_0$ na intervalu $\langle x_i, x_{i+1} \rangle$ a následně náhradě integrandu $f(x, y(x))$ vhodným interpolačním polynomem.

Tyto metody používají dvojici lineárních vícezkrokových metod. Jednou metodou se předpoví hodnota y_{i+1} a druhou metodou se získaná hodnota vylepší.

9. (1b.) Globální diskretizační chyba metody po i -tém kroku e_i je definována vztahem (lte_i značí lokální diskretizační chybu)

$$e_i = y(x_{i+1}) - y_i \qquad e_i = lte_0 + lte_1 + \dots + lte_i.$$

$$e_i = \max\{lte_0, lte_1, \dots, lte_i\}. \qquad e_i = y(x_i) - y_i.$$

10. (1b.) Jednokroková numerická metoda je konvergentní, jestliže
- | | |
|------------------------|------------------------|
| je řádu nejvýše jedna. | je řádu alespoň jedna. |
| je implicitní | je explicitní. |

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Kapitola 7

Numerické řešení parciálních diferenciálních rovnic

Cíle

Po prostudování této kapitoly budete schopni vysvětlit:

- jaké metody pro řešení PDR se používají,
- jaký je princip metody sítí,
- co se rozumí korektností, stabilitou, aproximací a konvergencí diferenčního schématu,

- jaké schéma se používá pro Poissonovu rovnici a jaké jsou podmínky konvergence,
- jaké schéma se používá pro rovnici vedení tepla a jaké jsou podmínky konvergence,
- jaké schéma se používá pro vlnovou rovnici a jaké jsou podmínky konvergence.

7.1 Klasifikace metod

Většinu parciálních diferenciálních rovnic nedokážeme řešit explicitně. Proto je nutné použít numerické metody. Jelikož tyto rovnice, často nazývané *rovnice matematické fyziky*, jsou základním nástrojem matematického modelování, je otázka jejich numerického řešení zásadní. Jde o složitou a rozsáhlou problematiku, jíž jsou věnovány stovky monografií a učebnic. Jejím podrobnější výklad by vyžadoval mnohem větší počet hodin, než který máme k dispozici. Proto se omezíme na ukázkou použití metody sítí, která je na pochopení nejjednodušší a má nejmenší nároky na znalost náročnějšího matematického aparátu. Tuto numerickou metodu demonstrujeme na třech jednoduchých typech rovnic, se kterými jsme se seznámili v předchozích přednáškách a které jsou časté a důležité v aplikacích.

Nicméně pro úplnost začneme přehledem nejvýznamnějších metod, které se pro

numerické řešení parciálních diferenciálních rovnic používají.

1. *Diferenční metoda* neboli *metoda sítí*. Jde o dnes již méně užívanou, ale pořád oblíbenou metodu.
2. *Metoda konečných objemů*. Používá se zejména při studiu proudění tekutin a plynů (hydrodynamika, mechanika kontinua).
3. *Variační metody* (Ritzova, Galerkinova, metoda nejmenších čtverců atd.). Používají se pro všechny typy rovnic, jsou vhodné i pro hledání tzv. slabých řešení.
4. *Metoda konečných prvků* (MKP). Jde rovněž o variační metodu, která má mimořádný význam v aplikacích, proto existuje řada speciálních programů pro řešení rovnic pomocí MKP. Používá se zejména na problémy mechaniky pevných těles (pružnost, pevnost). Viz [12]. K zakladatelům matematické teorie MKP patřil brněnský matematik [Miloš Zlámal](#) (1924–1997).

Při použití kterékoli metody nakonec dostáváme rozsáhlé soustavy lineárních algebraických rovnic (v případě nestacionárních rovnic v prostoru může jít o miliony neznámých). Matice těchto soustav mají často speciální vlastnosti (jsou řídké, pásové, reducibilní apod.) a na jejich řešení existuje řada speciálních metod. Ani touto problematikou se však

nemůžeme zabývat. Viz např. [17], [18], [49].

Jak již bylo řečeno, omezíme se na metodu sítí. Vysvětlíme její princip a ukážeme, jak se řeší některé lineární parciální diferenciální rovnice 2. řádu, kde neznámá funkce má dvě proměnné. Jde o rovnice tvaru

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + Fu = G(x, y). \quad (7.1)$$

Hledáme řešení $u(x, y)$ této rovnice, které je definované na množině Ω , přičemž jsou předepsány hodnoty na hranici $\partial\Omega$, tj. $u(x, y) = \varphi(x, y)$ pro $[x, y] \in \partial\Omega$, kde φ je daná (spojitá) funkce. Obdobně mohou být zadány na hranicích (nebo jejich částech) hodnoty parciálních derivací řešení $u(x, y)$ nebo jeho derivací ve směru normály k hranici, popř. hodnoty lineárních kombinací řešení a jeho derivací. Budeme předpokládat, že jde o tzv. klasické neboli silné řešení, tj. funkci, která je spojitá na Ω včetně její hranice a má potřebné spojitě derivace uvnitř této množiny.

Tyto rovnice se dělí na tři typy: eliptické, parabolické a hyperbolické. V dalším textu ukážeme, jak lze metodou sítí řešit jednoho zástupce každé třídy. Půjde postupně o Poissonovu resp. Laplaceovu rovnici, rovnici vedení tepla resp. difuzní rovnici a vlnovou rovnici.

7.2 Metoda sítí

Než se budeme věnovat řešení konkrétních parciálních diferenciálních rovnic metodou sítí, popíšeme její princip a stručně se zmíníme o některých důležitých pojmech a výsledcích, které s touto metodou souvisí.

7.2.1 Princip metody sítí

Sítí rozumíme konečnou množinu bodů, tzv. *uzlů*, které vhodně pokrývají množinu Ω , na níž parciální diferenciální rovnici vyšetřujeme. Nejčastěji se používá pravidelná čtvercová nebo obdélníková síť. Obecně však může být síť nepravidelná, tvořená např. vrcholy trojúhelníků, a vzdálenost mezi uzly proměnná. Řešení $u(x, y)$ najdeme pouze v uzlech a navíc jen přibližně.

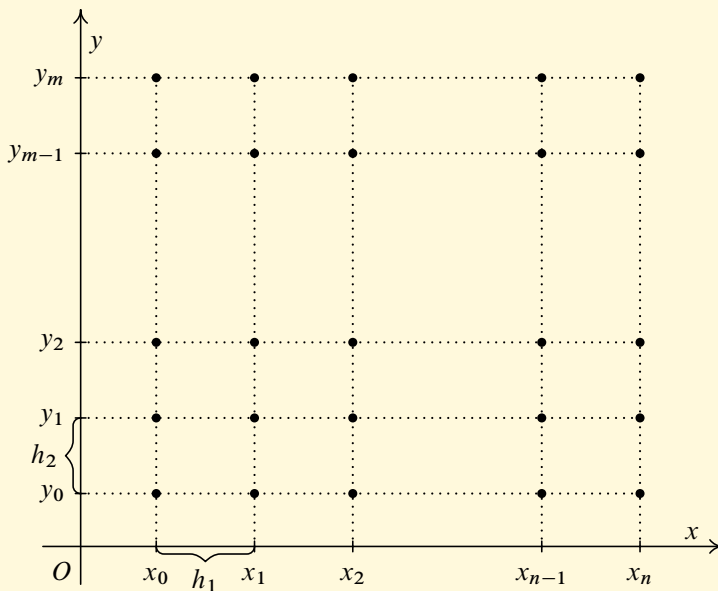
Uzly sítě dělíme na

- *hraniční* — tvoří *síťovou hranici*;
hraniční body leží na geometrické hranici $\partial\Omega$ nebo blízko ní (uvnitř i vně),
- *vnitřní* — leží i s vhodnými sousedy (tvořícími tzv. šablonu, viz dále) uvnitř Ω .

Protože ve vyšetřovaných úlohách budeme používat pouze obdélníkovou resp. čtvercovou síť, použijeme následující označení uzlů. Na ose x zvolíme ekvidistantní dělení s body x_0, \dots, x_n s krokem $h_1 > 0$ a na ose y zvolíme ekvidistantní dělení s body y_0, \dots, y_m s krokem $h_2 > 0$. Je tedy $x_1 = x_0 + h_1, x_2 = x_1 + h_1, \dots, x_n = x_{n-1} + h_1$ a obdobně $y_1 = y_0 + h_2, y_2 = y_1 + h_2, \dots, y_m = y_{m-1} + h_2$. Viz obr. 7.1. Uzly tudíž budou mít souřadnice $[x_i, y_j]$. Hodnoty řešení $u(x, y)$ v uzlech označíme u_{ij} , tj. $u_{ij} = u(x_i, y_j)$. V šabloně budeme pro názornost označovat uzel $[x_i, y_j]$ dvojicí i, j .

Naším cílem bude sestavit soustavu lineárních algebraických rovnic s neznámými v_{ij} , jejichž hodnoty budou aproximovat hodnoty u_{ij} , tj. přesné hodnoty řešení v uzlových bodech. Tuto soustavu budeme nazývat *síťové rovnice*. Při jejich vytváření budeme postupovat následujícím způsobem:

- 1) Zvolíme vhodnou skupinu několika blízkých bodů, která se nazývá *šablona*. Podoba šablony je specifická pro jednotlivé typy rovnic. Konkrétní příklady uvidíme dále. Šablona má obvykle jakýsi „ústřední“ uzel, který udává její polohu.



Obr. 7.1: Obdélníková síť

- 2) Parciální diferenciální rovnici na šabloně diskretizujeme. K tomu se používají různé postupy. V našem případě půjde vesměs o Taylorův vzorec vhodného řádu. Tak získáme vztah mezi hodnotami u_{ij} v uzlech šablony, který bude obsahovat neznámé veličiny odpovídající zbytkům z Taylorova vzorce. Tyto veličiny budou pro dostatečně jemnou síť malé. Jejich zanedbáním dostaneme přibližný lineární vztah mezi hodnotami u_{ij} . V něm symbol přibližné rovnosti \approx nahradíme symbolem rovnosti $=$, hodnoty u_{ij} nahradíme neznámými v_{ij} a obdržíme tzv. *diferenční schéma* neboli *síťovou rovnici*. Půjde tedy o náhradu derivací diferencemi.
- 3) Šablonu posunujeme po síti, umísťujeme ji do jednotlivých uzlů a pomocí diferenčního schématu vytváříme síťové rovnice. Přitom mohou nastat dva případy:
- Šablona je umístěna tak, že všechny její uzly jsou vnitřní.
Pak dostaneme bez potíží vztah mezi neznámými v_{ij} .
 - Šablona je umístěna tak, že alespoň jeden její bod je bodem síťové hranice.
Pak je nutné vhodným způsobem přenést okrajové a počáteční podmínky. Pokud je uzel přímo na geometrické hranici $\partial\Omega$ a je zde předepsána funkční hodnota, je hodnota v_{ij} odpovídající tomuto uzlu známá. Pokud není uzel přímo na geometrické hranici $\partial\Omega$ nebo je zde předepsána hodnota derivace řešení, je nutné použít nějakou aproximaci. Konkrétní postupy si ukážeme dále u jednotlivých typů rovnic.

7.2.2 Korektnost, stabilita, aproximace a konvergence

Aby hodnoty v_{ij} dávaly v uzlových bodech dobrou aproximaci přesných hodnot řešení u_{ij} , musí mít diferenční schéma vhodné vlastnosti. Těch si nyní všimneme. Protože jde o poměrně složitou problematiku, upustíme od přesných definic, náš výklad bude neformální a jeho cílem bude dát čtenáři představu, o jaké požadavky jde.

Označíme $\mathbf{h} = (h_1, h_2)$ a $\|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2}$ (eukleidovská norma). Tedy $\|\mathbf{h}\|$ je velikost úhlopříčky jednoho dílku uvažované obdélníkové sítě. Jak již bylo řečeno, pro určení neznámých v_{ij} dostaneme systém lineárních rovnic.

Označme $A_{\mathbf{h}}$ matici soustavy a $B_{\mathbf{h}}$ sloupec pravých stran tohoto systému. Dále nechť $V_{\mathbf{h}}$ je sloupec neznámých v_{ij} ; neznámé jsou seřazené nějakým pevně zvoleným vhodným způsobem, např. (viz obr. 7.1) podle uzlových bodů nejprve zleva doprava první řádek zdola, pak zleva doprava druhý řádek zdola atd., tj. pořadí bude $v_{00}, \dots, v_{n0}, v_{01}, \dots, v_{n1}, \dots, v_{0m}, \dots, v_{nm}$. Systém síťových rovnic potom můžeme zapsat pomocí součinu matic ve tvaru $A_{\mathbf{h}}V_{\mathbf{h}} = B_{\mathbf{h}}$. Index \mathbf{h} naznačuje, že systém závisí na volbách kroků sítě. Jak uvidíme později na konkrétních příkladech, (čtvercová) matice $A_{\mathbf{h}}$ je určena koeficienty rovnice (7.1), zatímco (sloupcová) matice $B_{\mathbf{h}}$ je určena pravou stranou této rovnice a počátečními a okrajovými podmínkami.

Nechť $\|\cdot\|$ značí nějakou vektorovou normu na \mathbb{R}^s , kde s značí počet řádků sloupce B_h . Na množině matic \mathbb{M}_s použijeme normu $\|\cdot\|$ s ní souhlasnou.

1) V oddílu 1.4 jsme definovali korektní úlohu. Tento požadavek je zvlášť významný u parciálních diferenciálních rovnic. Je proto přirozené požadovat, aby obdobnou vlastnost měly lineární algebraické rovnice, které vzniknou jejich diskretizací.

Řekneme, že diferenční schéma $A_h V_h = B_h$ je *korektní*, jestliže pro dostatečně malé $\|h\|$ má pro libovolnou pravou stranu B_h *jediné řešení* a je *stabilní*. Přitom stabilitou rozumíme spojitou závislost řešení V_h na pravé straně B_h stejnoměrně vzhledem k h .

Vysvětlíme tyto pojmy podrobněji. Z lineární algebry víme, že soustava lineárních rovnic $A_h V_h = B_h$ může mít teoreticky jediné řešení, nekonečně mnoho řešení nebo nemá žádné řešení. Z požadavku jednoznačnosti pro libovolnou pravou stranu B_h vyplývá, že matice soustavy A_h musí být čtvercová s nenulovým determinanem. To znamená, že je regulární, a existuje tudíž inverzní matice A_h^{-1} . Řešení V_h je tudíž dáno vzorcem $V_h = A_h^{-1} B_h$, takže platí $\|V_h\| \leq \|A_h^{-1}\| \cdot \|B_h\|$. Požadavek stability pak značí, že existuje konstanta $M > 0$ (nezávislá na $\|h\|$) taková, že $\|A_h^{-1}\| \leq M$ pro libovolné dostatečně malé $\|h\|$. Platí tedy, že $\|V_h\| \leq M \cdot \|B_h\|$. Tato vlastnost

obecně závisí na volbě normy v \mathbb{R}^s a jí indukované maticové normy. Může se stát, že pro některou normu je splněna a pro jinou ne.

- 2) Označme $U_{\mathbf{h}}$ sloupec obsahující přesné hodnoty u_{ij} a položme $W_{\mathbf{h}} = V_{\mathbf{h}} - U_{\mathbf{h}}$. Sloupec $W_{\mathbf{h}}$ vyjadřuje chybu aproximace přesného řešení v uzlových bodech pomocí řešení diferenčního schématu $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$. Dosazením do tohoto schématu dostaneme pro chybu vztah $A_{\mathbf{h}} W_{\mathbf{h}} = B_{\mathbf{h}} - A_{\mathbf{h}} U_{\mathbf{h}}$.

Řekneme, že diferenční schéma $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$ *aproximuje* úlohu (7.1), jestliže platí $\|B_{\mathbf{h}} - A_{\mathbf{h}} U_{\mathbf{h}}\| \rightarrow 0$ pro $\|\mathbf{h}\| \rightarrow 0$.

Řekneme, že diferenční schéma $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$ *aproximuje* úlohu (7.1) *s řádem* k , $k > 0$, jestliže $\|B_{\mathbf{h}} - A_{\mathbf{h}} U_{\mathbf{h}}\| = O(\|\mathbf{h}\|^k)$.

- 3) Diferenční schéma $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$ se nazývá *konvergentní*, jestliže pro něj platí vztah $\|V_{\mathbf{h}} - U_{\mathbf{h}}\| \rightarrow 0$ pro $\|\mathbf{h}\| \rightarrow 0$.

Říkáme, že diferenční schéma $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$ *konverguje s řádem* k , $k > 0$, jestliže $\|V_{\mathbf{h}} - U_{\mathbf{h}}\| = O(\|\mathbf{h}\|^k)$.

Konvergence diferenčního schématu je důležitá, protože zajišťuje, že při dostatečně jemné síti jsou hodnoty v_{ij} dobrými aproximacemi přesného řešení u_{ij} v uzlových bodech.

Čím je řád konvergence vyšší, tím je přesnost lepší. Proto je důležitý následující výsledek (přesná formulace a důkaz viz např. [47, str. 96] nebo [57, str. 289]):

Schéma, které je stabilní a aproximuje úlohu (7.1), je konvergentní. Přitom je-li aproximace řádu k , je i konvergence řádu k .

7.2.3 Náhrada derivací diferencemi

V rovnicích, které budeme dále vyšetřovat, bude třeba na šabloně nahradit první a druhé nesmíšené parciální derivace řešení. Připravíme si potřebné vzorce.

Připomeňme, že pro funkci $f(x)$, která má v bodě x n -tou derivaci, $n \in \mathbb{N}$, platí Taylorův vzorec

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2!}f''(x)h^2 + \dots + \frac{1}{n!}f^{(n)}(x)h^n + R_n(x+h),$$

kde h je v absolutní hodnotě malé číslo a $R_n(x+h)$ je tzv. zbytek. Je-li derivace řádu $n+1$ spojitá v okolí bodu x , lze zbytek vyjádřit v Lagrangeově tvaru

$$R_n(x+h) = \frac{1}{(n+1)!}f^{(n+1)}(x+\theta h)h^{n+1}, \quad 0 < \theta < 1.$$

Z tohoto vyjádření je vidět, že platí $R_n(x + h) = O(h^{n+1})$.

Pro $n = 1$ má vzorec podobu

$$f(x + h) = f(x) + f'(x)h + O(h^2),$$

takže po osamostatnění $f'(x)$ vyjde

$$f'(x) = \frac{f(x + h) - f(x)}{h} + O(h). \quad (7.2)$$

Pro $n = 3$ použijeme vzorec dvakrát: pro $x + h$ a $x - h$:

$$f(x + h) = f(x) + f'(x)h + \frac{1}{2!}f''(x)h^2 + \frac{1}{3!}f'''(x)h^3 + O(h^4),$$

$$f(x - h) = f(x) - f'(x)h + \frac{1}{2!}f''(x)h^2 - \frac{1}{3!}f'''(x)h^3 + O(h^4).$$

Po sečtení a osamostatnění $f''(x)$ dostaneme

$$f''(x) = \frac{f(x + h) - 2f(x) + f(x - h)}{h^2} + O(h^2). \quad (7.3)$$

Pro náhrady budeme používat tudíž vzorce

$$\begin{aligned}f'(x) &\approx \frac{f(x+h) - f(x)}{h}, \\f''(x) &\approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2},\end{aligned}\tag{7.4}$$

přičemž v prvním případě je chyba řádu $O(h)$ a ve druhém řádu $O(h^2)$.

Všimněte si, že čitatele lze vyjádřit pomocí diferencí vpřed: $f(x+h) - f(x) = \Delta f(x)$, $f(x+h) - 2f(x) + f(x-h) = \Delta^2 f(x-h)$. Jde tedy opravdu o náhradu derivací diferencemi.

Kdybychom ponechali zbytky v Lagrangeově tvaru, mohli bychom nalézt přesné výrazy pro chyby, které by se shodovaly s (5.52) a (5.58).

7.3 Dirichletova úloha pro Poissonovu rovnici

Uvažujme rovnici

$$u_{xx} + u_{yy} = f(x, y), \quad (x, y) \in \Omega, \quad (7.5)$$

s okrajovou podmínkou

$$u(x, y) = \varphi(x, y), \quad (x, y) \in h(\Omega), \quad (7.6)$$

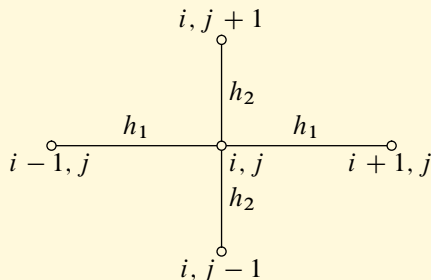
kde $f(x, y)$ resp. $\varphi(x, y)$ jsou dané spojitě funkce, definované na „rozumné“ ohraničené a uzavřené množině Ω resp. její hranici $\partial\Omega$. Tato rovnice je eliptického typu a nazývá se *Poissonova*¹ resp. pro $f(x, y) \equiv 0$ *Laplaceova*². Úloha, kdy máme najít řešení této rovnice s předepsanými hodnotami na hranici množiny Ω , se nazývá *Dirichletova*³.

¹**Siméon Denis Poisson** (1781–1840) (čti puason) — významný francouzský matematik. Zabýval se diferenciálními rovnicemi, pravděpodobností, pružností apod. Říkával často: „Život je dobrý jen na dvě věci: zabývání se matematikou a její vyučování.“

²**Pierre-Simon Laplace** (1749–1827) (čti laplas) — významný francouzský matematik, fyzik a astronom. Zabýval se parciálními diferenciálními rovnicemi a teorií pravděpodobnosti.

³**Johann Peter Gustav Lejeune Dirichlet** (1805–1859) (čti dirikle) — významný německý matematik. Zabýval se teorií čísel, matematickou analýzou a rovnicemi matematické fyziky.

Zvolíme obdélníkovou síť s kroky $h_1 > 0$, $h_2 > 0$ a použijeme následující šablonu:



Odvodíme nyní náhradu Poissonovy rovnice na této šabloně. Označíme $f_{ij} = f(x_i, y_j)$. Druhé derivace v Laplaceově operátoru $\Delta u = u_{xx} + u_{yy}$ ve vnitřních uzlech sítě nahradíme pomocí vzorce (7.3). Vyjde:

$$\frac{u_{i-1,j} - 2u_{ij} + u_{i+1,j}}{h_1^2} + O(h_1^2) + \frac{u_{i,j-1} - 2u_{ij} + u_{i,j+1}}{h_2^2} + O(h_2^2) = f_{ij}.$$

Zanedbáním členů $O(h_1^2)$ a $O(h_2^2)$ dostaneme pro přibližné hodnoty řešení v_{ij} sítovou

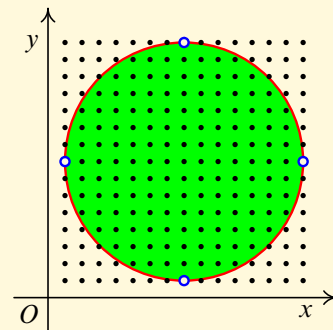
rovnici neboli diferenční schéma

$$\frac{v_{i-1,j} - 2v_{ij} + v_{i+1,j}}{h_1^2} + \frac{v_{i,j-1} - 2v_{ij} + v_{i,j+1}}{h_2^2} = f_{ij}. \quad (7.7)$$

V případě čtvercové sítě, tj. $h_1 = h_2 = h$, bude tato rovnice ještě jednodušší:

$$\frac{v_{i-1,j} + v_{i+1,j} + v_{i,j-1} + v_{i,j+1} - 4v_{ij}}{h^2} = f_{ij}. \quad (7.8)$$

Pomocí vzorce (7.7) můžeme sestavit rovnici pro každý uzel i, j , pro nějž všechny body šablony leží uvnitř množiny Ω nebo přesně na hranici $\partial\Omega$ (v bodě, který je přesně na hranici, je hodnota $u(x, y)$ známá, je daná funkcí $\varphi(x, y)$, protože musí být splněna okrajová podmínka (7.6)). Pokud ale některý bod šablony padne mimo množinu Ω , musíme příslušnou rovnici sestavit jinak. Situaci ilustruje obr. 7.2, kde Ω je kruh — přesně na hraniční kružnici jsou pouze čtyři body označené kroužkem.



Obr. 7.2: Dirichletova úloha na kruhu

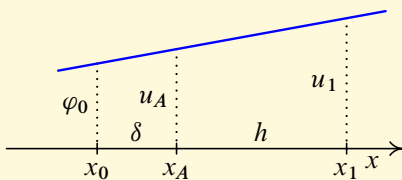
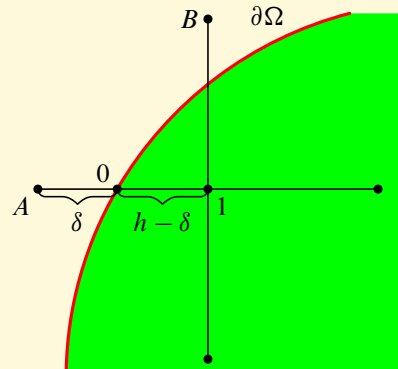
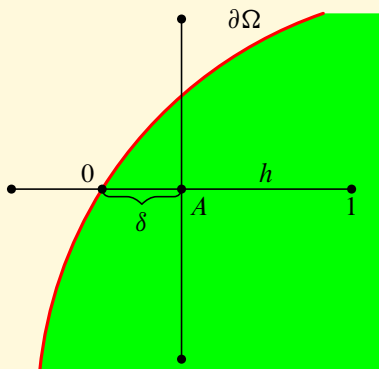
Přenos okrajových podmínek

Je-li uzel velmi blízko hranice, je (jako hrubou náhradou) možné vzít za hodnotu v tomto uzlu hodnotu funkce $\varphi(x, y)$ v nejbližším hraničním bodě. Jinou možností je provést lineární interpolaci nebo extrapolaci pomocí hodnot v blízkých bodech. Ukážeme, jak je možné postupovat (viz např. [57, str. 258]).

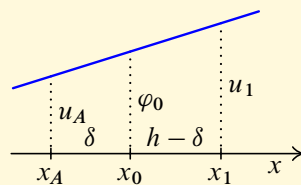
Použijeme označení z obr. 7.3. Dva body označíme indexy 0 a 1 a třetí písmenem A . Jejich souřadnice nechť jsou po řadě $[x_0, y_0]$, $[x_1, y_1]$ a $[x_A, y_A]$. Vzdálenost bodu A a bodu s indexem 0 je δ . Konečně označíme $u(x_A, y_A) = u_A$, $u(x_1, y_1) = u_1$ a $u(x_0, y_0) = \varphi(x_0, y_0) = \varphi_0$ (z okrajové podmínky).

Funkci $u(x, y)$ nahradíme nad přímkou $y = y_A$ lineární funkcí, která prochází body s indexy 0 a 1. Jinými slovy, funkci $u(x, y_A)$ jedné proměnné nahradíme Lagrangeovým polynomem určeným dvojicí $[x_0, \varphi_0]$ a $[x_1, u_1]$. Pak bude platit

$$u(x, y_A) \approx \varphi_0 \frac{x - x_1}{x_0 - x_1} + u_1 \frac{x - x_0}{x_1 - x_0}.$$



a) Interpolace



b) Extrapolace

Obr. 7.3: Přenos okrajových podmínek

Po dosazení $x = x_A$ dostaneme:

$$\text{Interpolace: } u(x_A, y_A) \approx \varphi_0 \frac{-h}{-(\delta + h)} + u_1 \frac{\delta}{\delta + h} = \frac{h\varphi_0 + \delta u_1}{h + \delta},$$

$$\text{Extrapolace: } u(x_A, y_A) \approx \varphi_0 \frac{-h}{-(h - \delta)} + u_1 \frac{-\delta}{h - \delta} = \frac{h\varphi_0 - \delta u_1}{h - \delta}.$$

V prvním případě přímo získáme síťovou rovnici odpovídající uzlu A . V druhém případě dostaneme síťovou rovnici pro přidaný uzel A (neleží v Ω). Obdobně vytvoříme extrapolací síťovou rovnici pro přidaný uzel B (viz obr. 7.3 b)) a pak už můžeme použít základní šablonu umístěnou do uzlu $[x_1, y_1]$.

Existují i jiné, lepší způsoby přenosu okrajových podmínek, viz např. [47, str. 218] nebo [57, str. 253].

U složitější hranice bude uzlů, v nichž je třeba přenést okrajovou podmínku, velký počet. Je nemožné sestavovat příslušné rovnice ručně a je třeba vytvořit vhodný algoritmus pro jejich automatické vygenerování. To však může být velmi komplikované.

Výsledkem bude soustava lineárních algebraických rovnic o velkém počtu neznámých (u úloh z praxe i stovky tisíc neznámých). Matice soustavy bude řídká (bude obsahovat

velký počet nul) a ireducibilní (soustava se nerozpadne na menší soustavy). Takové soustavy se obvykle řeší iteračními metodami (Jacobiova, Gaussova-Seidlova, metoda sdružených gradientů apod.).

Lze dokázat (viz [47, str. 232]), že pro dostatečně hladké funkce $f(x, y)$ a $\varphi(x, y)$ je popsané diferenční schéma *stabilní* a *aproximuje* úlohu (7.5) – (7.6) s řádem 2, takže je rovněž *konvergentní* s řádem 2., Tedy pro $\|\mathbf{h}\| \rightarrow 0$ hodnoty v uzlových bodech konvergují k hodnotě přesného řešení a chyba je $O(\|\mathbf{h}\|^2)$.

Ukázka 7.1 Uvažujme Poissonovu rovnici

$$u_{xx} + u_{yy} = 30x^2 - y^2, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1,$$

s Dirichletovými okrajovými podmínkami

$$u(x, 0) = x, \quad 0 \leq x \leq 1,$$

$$u(x, 1) = 0,5x(x - 1) + 1, \quad 0 \leq x \leq 1,$$

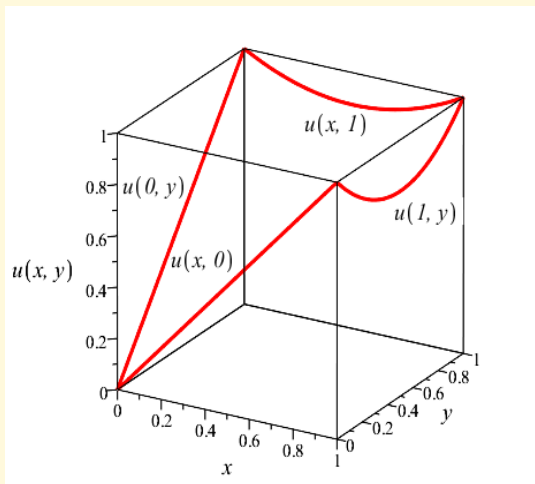
$$u(0, y) = y, \quad 0 \leq y \leq 1,$$

$$u(1, y) = 0,8y(y - 1) + 1, \quad 0 \leq y \leq 1.$$

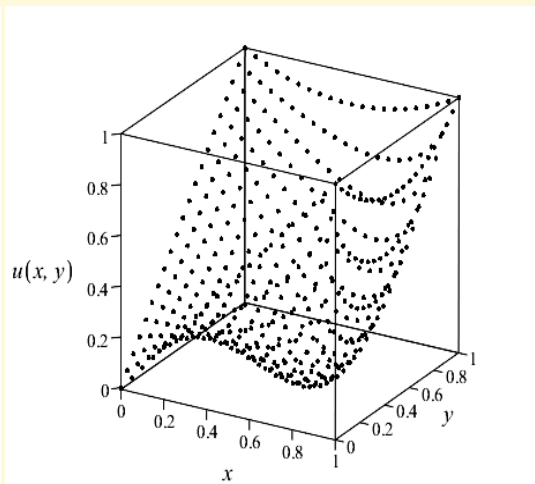
Množina Ω je uzavřený čtverec $\langle 0, 1 \rangle \times \langle 0, 1 \rangle$, jehož hranice $h(\Omega)$ je tvořena čtyřmi úsečkami. Okrajové podmínky jsou zadané tak, že funkce $\varphi(x, y)$ je na této hranici spojitá.

Při řešení byla použita čtvercová síť s krokem $h = 1/20$. Počet bodů sítě tedy je $21^2 = 441$. Odečteme-li body na hranici, v nichž známe přesné hodnoty řešení, dostaneme čtvercovou soustavu lineárních rovnic o $19^2 = 361$ neznámých. Vnitřním uzlům odpovídají rovnice o pěti neznámých, hraničním uzlům (tj. takovým, kdy některý bod šablony bude ležet na geometrické hranici obdélníka) rovnice o třech nebo čtyřech

neznámých. Řešením této soustavy dostaneme přibližné hodnoty v_{ij} v uzlových bodech. Na obr. 7.4 a) jsou zobrazeny okrajové podmínky a na obr. 7.4 b) jsou zobrazené přibližné hodnoty řešení $u(x, y)$ této úlohy v uzlových bodech. Na obr. 7.5 je těmito body proložena plocha dávající představu o řešení na celém čtverci.

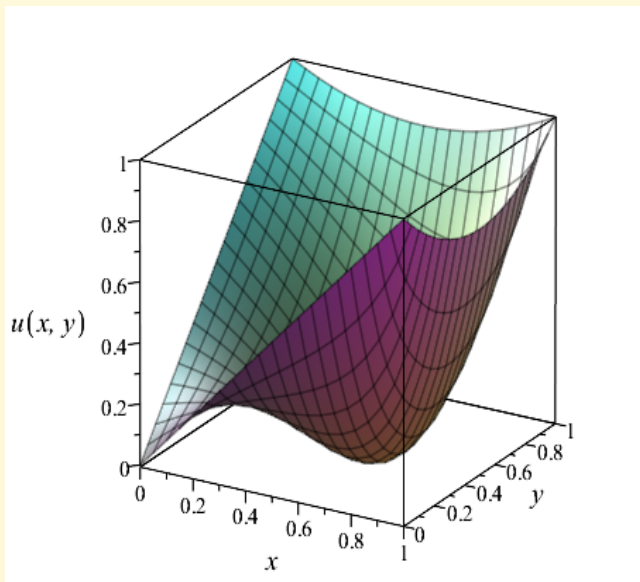


a) Okrajové podmínky



b) Hodnoty řešení v uzlových bodech

Obr. 7.4: Řešení Poissonovy rovnice



Obr. 7.5: Graf řešení Poissonovy rovnice

7.4 Smíšená úloha pro rovnici vedení tepla

Uvažujme rovnici difuzního typu

$$u_t = a^2 u_{xx} + f(x, t), \quad (7.9)$$

kde $u = u(x, t)$, $0 \leq x \leq l$, $t \geq 0$, $l > 0$, $a > 0$, s počáteční podmínkou

$$u(x, 0) = \psi(x), \quad 0 \leq x \leq l, \quad (7.10)$$

a okrajovými podmínkami

$$u(0, t) = \varphi_1(t), \quad u(l, t) = \varphi_2(t), \quad t \geq 0, \quad (7.11)$$

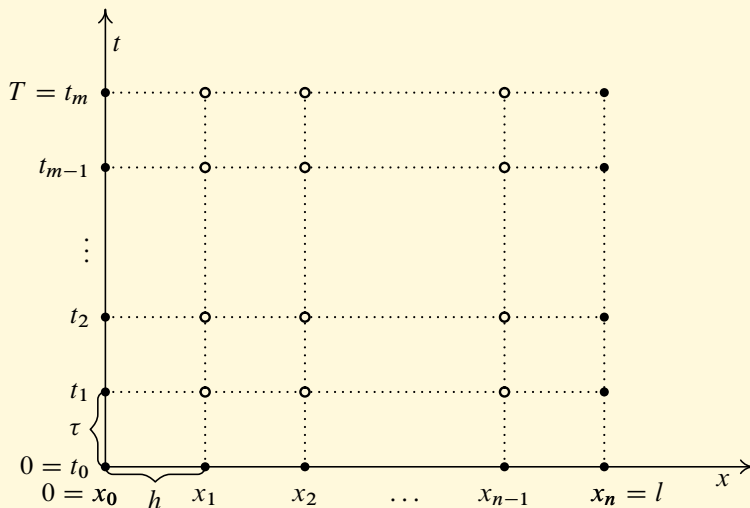
kde $\psi(x)$, $\varphi_1(t)$ a $\varphi_2(t)$ jsou spojité funkce takové, že $\psi(0) = \varphi_1(0)$ a $\psi(l) = \varphi_2(0)$.

Tato rovnice může popisovat např. vedení tepla v tyči délky l . Funkce $u(x, t)$ udává teplotu tyče v bodě x v čase t . Funkce $\psi(x)$ popisuje rozložení teploty na počátku, tj. v čase $t = 0$, a funkce $\varphi_1(t)$ a $\varphi_2(t)$ popisují teploty na koncích tyče. Funkce $f(x, t)$ popisuje rozložení (hustotu) tepelných zdrojů. Jiný jev, který tato rovnice může popisovat, je difuze

v tenké trubici. Pak $u(x, t)$ udává koncentraci difundující látky v bodě x v čase t , funkce $\varphi(x)$ popisuje její počáteční koncentraci, funkce $\varphi_1(t)$ a $\varphi_2(t)$ popisují koncentrace na koncích trubice a funkce $f(x, t)$ představuje zdroje.

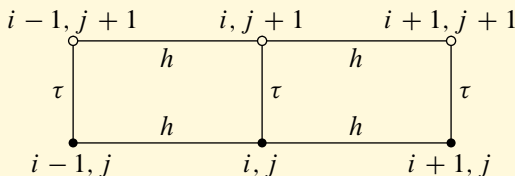
Řešení nemůžeme hledat na nekonečném časovém intervalu, proto se omezíme na interval $t \in \langle 0, T \rangle$, kde $T > 0$. Volíme obdélníkovou síť s kroky $h > 0$, $\tau > 0$, kde $h = l/n$, $\tau = T/m$, $m, n \in \mathbb{N}$. Tedy interval $\langle 0, l \rangle$ rozdělíme na n stejných dílků a interval $\langle 0, T \rangle$ na m stejných dílků — viz obr. 7.6. Označíme $u(x_i, t_j) = u_{ij}$, $i = 0, \dots, n$, $j = 0, \dots, m$.

Hodnoty v hraničních uzlech známe: $u_{i0} = \psi(x_i)$, $i = 0, \dots, n$, a $u_{0j} = \varphi_1(t_j)$, $u_{nj} = \varphi_2(t_j)$, $j = 0, \dots, m$. Na obr. 7.6 jsou tyto uzly vyznačeny plnými kroužky. Ve zbývajících uzlech (jsou vyznačeny prázdnými kroužky) musíme hodnoty řešení určit.



Obr. 7.6: Rovnice vedení tepla — volba sítě

Pro sestavení síťových rovnic použijeme následující dvouvrstvou šestibodovou šablonu:



V rovnici (7.9) nahradíme pomocí vztahů (7.2) a (7.3) derivace. Nahrazení u_t bude přímočaré:

$$u_t(x_i, t_j) = \frac{u_{i,j+1} - u_{ij}}{\tau} + O(\tau).$$

V případě u_{xx} budeme postupovat trochu složitěji. Platí

$$u_{xx}(x_i, t_j) = \frac{u_{i-1,j} - 2u_{ij} + u_{i+1,j}}{h^2} + O(h^2),$$

$$u_{xx}(x_i, t_{j+1}) = \frac{u_{i-1,j+1} - 2u_{i,j+1} + u_{i+1,j+1}}{h^2} + O(h^2).$$

Nyní zvolíme reálné číslo σ a z posledních dvou vztahů vytvoříme vážený průměr s vahami σ a $1 - \sigma$, kterým aproximujeme druhou derivaci u_{xx} . Konečně označíme $f_{ij} = f(x_i, t_j + 0,5\tau)$ (bod $[x_i, t_j + 0,5\tau]$ leží uprostřed mezi body $[x_i, t_j]$ a $[x_i, t_{j+1}]$). Po zanedbání zbytků dostaneme pro přibližné hodnoty řešení v_{ij} vztah

$$\frac{v_{i,j+1} - v_{ij}}{\tau} = a^2 \left[\sigma \frac{v_{i-1,j+1} - 2v_{i,j+1} + v_{i+1,j+1}}{h^2} + (1 - \sigma) \frac{v_{i-1,j} - 2v_{ij} + v_{i+1,j}}{h^2} \right] + f_{ij},$$

z něhož po úpravě obdržíme tzv. *diferenční schéma s vahou*

$$\begin{aligned} v_{i,j+1} - \sigma \frac{a^2 \tau}{h^2} (v_{i-1,j+1} - 2v_{i,j+1} + v_{i+1,j+1}) &= \\ &= v_{ij} + (1 - \sigma) \frac{a^2 \tau}{h^2} (v_{i-1,j} - 2v_{ij} + v_{i+1,j}) + \tau f_{ij}. \end{aligned} \quad (7.12)$$

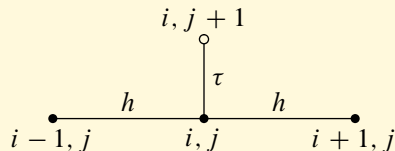
Tato lineární síťová rovnice váže přibližné hodnoty řešení v šesti bodech výše uvedené šablony. Všimneme si tří speciálních případů.

Explicitní schéma ($\sigma = 0$)

Diferenční schéma bude mít následující podobu:

$$v_{i,j+1} = v_{ij} + \frac{a^2\tau}{h^2} (v_{i-1,j} - 2v_{ij} + v_{i+1,j}) + \tau f_{ij}. \quad (7.13)$$

Levá strana obsahuje hodnotu pouze v jednom bodě ($j + 1$)-ní vrstvy, takže šablona se zjednoduší:



Protože známe hodnoty řešení v_{ij} v nulté vrstvě, tj. pro $j = 0$, a na levé a pravé straně obdélníku $\langle 0, l \rangle \times \langle 0, T \rangle$ — na obr. 7.6 jsou tyto uzly vyznačeny plnými kroužky, umožňuje nám vzorec (7.13) počítat řešení po vrstvách. Stačí umístit základnu šablony na nultou vrstvu, tj. ve vzorci (7.13) zvolit $j = 0$, a např. odleva doprava postupně určit hodnoty neznámých $v_{11}, \dots, v_{n-1,1}$. Pak šablonu zvedneme o jednu vrstvu, tj. zvolíme $j = 1$,

a obdobně určíme hodnoty neznámých $v_{12}, \dots, v_{n-1,2}$. A tak dále. Nevzniká tedy rozsáhlá soustava lineárních rovnic, kterou je nutné řešit jako celek. Proto se toto schéma nazývá *explicitní*.

Je-li $\sigma \neq 0$, nazývá se schéma *implicitní* a situace je jiná. Levá strana vzorce (7.12) obsahuje s nenulovými koeficienty přibližné hodnoty řešení ve třech sousedních bodech $(j+1)$ -ní vrstvy. Pro $j = 0$ můžeme dosadit do pravé strany známé hodnoty v nulté vrstvě. Posouváme-li šablonu postupně zleva doprava, tj. volíme-li $i = 1, \dots, n-1$, dostaneme soustavu $n-1$ lineárních rovnic pro neznámé $v_{11}, \dots, v_{n-1,1}$ v první vrstvě. Přitom první rovnice obsahuje pouze neznámé v_{11} a v_{21} (hodnotu v_{01} známe přesně), druhá rovnice obsahuje neznámé v_{11}, v_{21} a v_{31} atd. Předposlední rovnice obsahuje neznámé $v_{n-3,1}, v_{n-2,1}$ a $v_{n-1,1}$ a poslední pouze neznámé $v_{n-2,1}$ a $v_{n-1,1}$ (hodnotu v_{n1} opět známe přesně). Matice soustavy je tudíž třídiagonální a na řešení takových soustav existují velmi účinné metody.

Podobně postupujeme dále. Ve vzorci (7.12) zvolíme $j = 1$ a sestavíme stejným způsobem soustavu lineárních rovnic pro neznámé $v_{12}, \dots, v_{n-1,2}$. A tak dále. Na rozdíl od eliptického případu (Laplaceova a Poissonova rovnice) nemusíme řešit jednu obrovskou soustavu pro všechny neznámé v_{ij} jako celek, dojde k rozpadnutí na samostatné soustavy

pro každou vrstvu zvlášť.

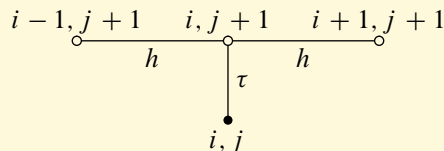
V praxi se používají zejména následující dva případy implicitních schémat.

Ryze implicitní schéma ($\sigma = 1$)

Diferenční schéma bude mít následující podobu:

$$v_{i,j+1} - \frac{a^2 \tau}{h^2} (v_{i-1,j+1} - 2v_{i,j+1} + v_{i+1,j+1}) = v_{ij} + \tau f_{ij}. \quad (7.14)$$

Pravá strana obsahuje hodnotu pouze v jednom bodě j -té vrstvy, takže šablona se zjednoduší:



Jedná se o implicitní schéma, jehož použití bylo popsáno výše. Postupujeme odspodu a pro každou vrstvu musíme řešit soustavu lineárních rovnic s třídiagonální maticí. Pro-

tože z předchozí vrstvy se používá vždy jen jedna hodnota, nazývá se toto schéma *ryze implicitní*.

Crankovo-Nicolsonové schéma ($\sigma = 1/2$)

Diferenční schéma bude mít následující podobu:

$$\begin{aligned} v_{i,j+1} - \frac{a^2\tau}{2h^2}(v_{i-1,j+1} - 2v_{i,j+1} + v_{i+1,j+1}) &= \\ &= v_{ij} + \frac{a^2\tau}{2h^2}(v_{i-1,j} - 2v_{ij} + v_{i+1,j}) + \tau f_{ij}. \end{aligned} \quad (7.15)$$

Používá se tedy obecná dvouvrstvá šestibodová šablona. Jedná se o implicitní schéma, jehož použití bylo popsáno výše. Postupujeme odspodu a pro každou vrstvu musíme řešit soustavu lineárních rovnic s třídiagonální maticí. Toto schéma se nazývá *Crankovo*¹-*Nicolsonové*². Níže uvidíme, proč je právě volba $\sigma = 1/2$ zajímavá.

¹**John Crank** (1916–2006) (čti krenk) — anglický numerický matematik.

²**Phyllis Nicolsonová** (1917–1968) (čti nikolsnová) — anglická fyzička.

Lze dokázat (viz [46, str. 238] nebo [47, str. 262]), že pro dostatečně hladké funkce $f(x, t)$, $\psi(x)$, $\varphi_1(t)$ a $\varphi_2(t)$ diferenční schéma (7.12) *aproximuje* úlohu (7.9)–(7.11) s přesností $O(h^2 + \tau)$ pro $\sigma \neq 1/2$ a s přesností $O(h^2 + \tau^2)$ pro $\sigma = 1/2$ (Crankovo-Nicolsonové schéma).

Dále lze ukázat (viz [46, str. 240] nebo [47, str. 268]), že pokud je $\sigma \geq 0$ a je splněna podmínka

$$\sigma \geq \frac{1}{2} - \frac{h^2}{4a^2\tau}, \quad (7.16)$$

je schéma (7.12) *stabilní* (v jisté normě).

Tedy v případě *explicitního schématu* (7.13) je stabilita zajištěna podmínkou

$$\frac{a^2\tau}{h^2} \leq \frac{1}{2} \quad (7.17)$$

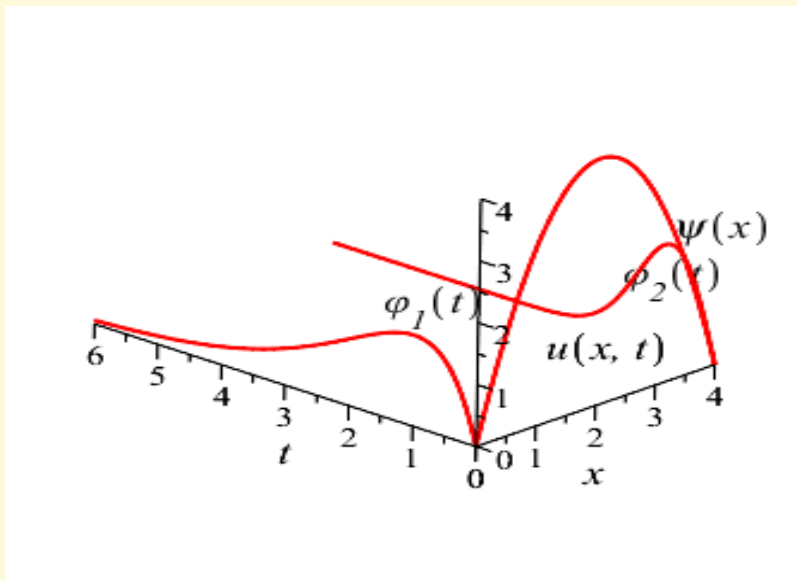
(říkáme, že toto schéma je neabsolutně stabilní). Schéma je konvergentní s řádem jedna (chyba je $O(\tau + h^2)$). Podmínka (7.17) je podstatná. Pokud není splněna, vychází zcela nesmyslné výsledky. Podmínka vynucuje, že krok τ musí být vzhledem k h poměrně velmi malý, protože musí platit $\tau \leq h^2/(2a^2)$.

Při označení z (7.9) platí, že $a = 1$ a $l = 4$. Dále je $\psi(x) = x(4 - x)$, $\varphi_1(t) = 4te^{-t}$, $\varphi_2(t) = 4te^{-t^2}$ a $f(x, t) = 0$. Rovnice může popisovat např. vedení tepla v tyči.

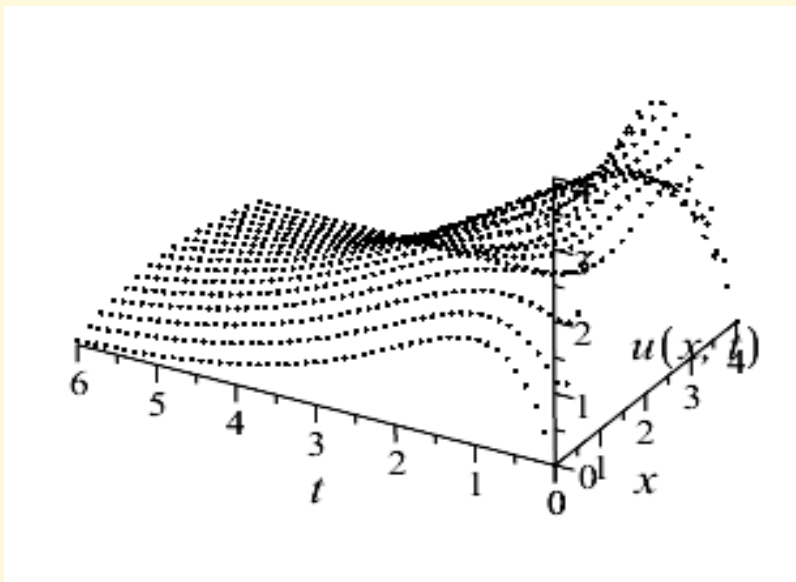
Použijeme Crankovo-Nicolsonové implicitní schéma (7.15). Zvolili jsme $T = 6$, řešení tedy bylo nalezeno na obdélníku $\langle 0, 4 \rangle \times \langle 0, 6 \rangle$. Dále jsme vybrali $n = 15$, $m = 50$, takže $h = 4/15 \doteq 0,27$ a $\tau = 6/50 = 0,12$. Použité schéma je vždy konvergentní. Počet bodů sítě je $16 \cdot 51 = 816$. Odečteme-li hraniční body, v nichž známe přesné hodnoty řešení, zůstane 700 neznámých. Vzniklou soustavu síťových rovnic nemusíme řešit jako celek, rozpadne se na 50 menších soustav odpovídajících neznámým v jednotlivých vrstvách, které mají po 14 neznámých. Neznámé udávají přibližné hodnoty v_{ij} řešení v uzlových bodech.

Na obr. 7.7 jsou zobrazeny okrajové podmínky a počáteční podmínka a na obr. 7.8 jsou zobrazené přibližné hodnoty řešení $u(x, t)$ této úlohy v uzlových bodech. Na obr. 7.9 je těmito body proložena plocha dávající představu o řešení na celém obdélníku. Na obr. 7.10 je animace znázorňující vyrovnávání teploty tyče v časovém intervalu $\langle 0, 6 \rangle$.

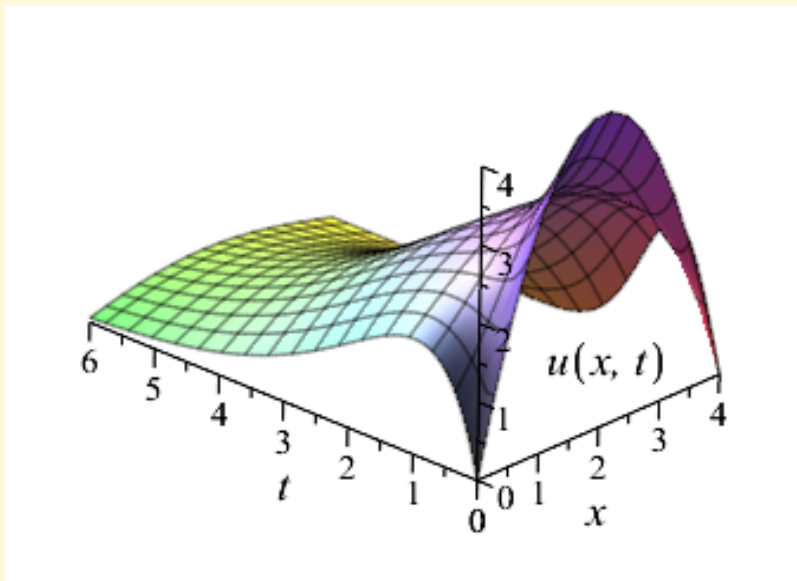
Na obr. 7.11 je zobrazen graf řešení, které bylo získáno na téže síti pomocí schématu (7.12) pro $\sigma = 0,31$. Pro tuto hodnotu není splněna podmínka stability (7.16) — vyjde $1/2 - h^2/(4a^2\tau) \doteq 0,352 > \sigma$. Z obrázku je patrné, že výsledek je zcela nesmyslný. Stabilita schématu je tudíž podstatná vlastnost pro jeho konvergenci.



Obr. 7.7: Řešení rovnice vedení tepla — okrajové a počáteční podmínky

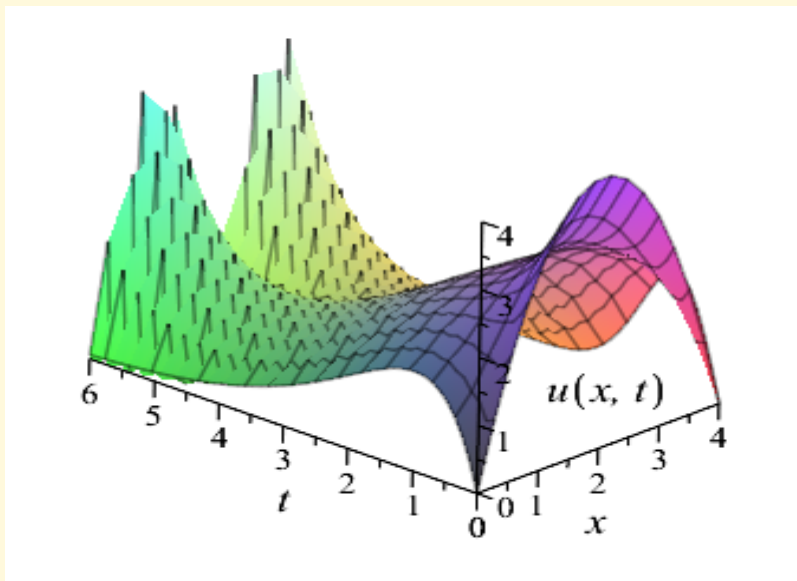


Obr. 7.8: Řešení rovnice vedení tepla — hodnoty řešení v uzlových bodech



Obr. 7.9: Graf řešení rovnice vedení tepla

Obr. 7.10: Animace řešení rovnice vedení tepla (vyrovnávání teploty tyče)



Obr. 7.11: Graf řešení rovnice vedení tepla získaného nestabilní metodou

7.5 Smíšená úloha pro vlnovou rovnici

Uvažujme vlnovou rovnici

$$u_{tt} = a^2 u_{xx} + f(x, t), \quad (7.18)$$

kde $u = u(x, t)$, $0 \leq x \leq l$, $t \geq 0$, $l > 0$, $a > 0$, s počátečními podmínkami

$$u(x, 0) = \psi_1(x), \quad u_t(x, 0) = \psi_2(x), \quad 0 \leq x \leq l, \quad (7.19)$$

a okrajovými podmínkami

$$u(0, t) = \varphi_1(t), \quad u(l, t) = \varphi_2(t), \quad t \geq 0, \quad (7.20)$$

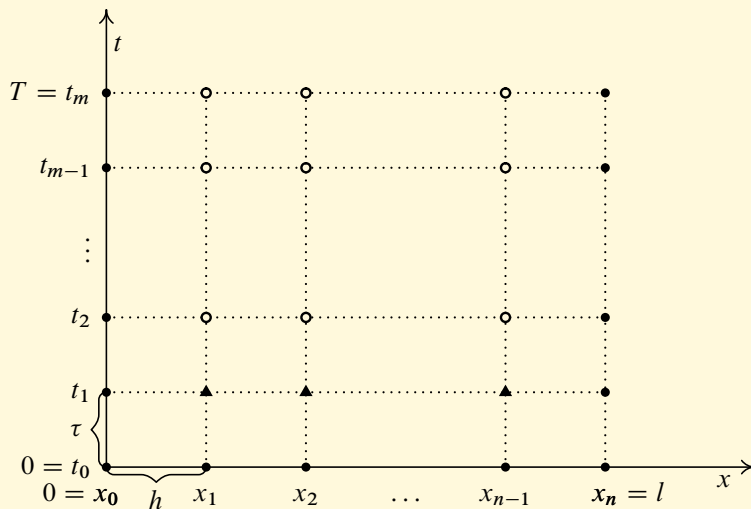
kde $\psi_1(x)$, $\psi_2(x)$, $\varphi_1(t)$ a $\varphi_2(t)$ jsou spojité funkce takové, že $\psi_1(0) = \varphi_1(0)$, $\psi_1(l) = \varphi_2(0)$, $\psi_2(0) = \varphi_1'(0)$ a $\psi_2(l) = \varphi_2'(0)$.

Tato rovnice může popisovat např. kmity struny délky l . Funkce $u(x, t)$ udává polohu struny v bodě x v čase t . Funkce $\psi_1(x)$ popisuje počáteční tvar struny, tj. v čase $t = 0$, a funkce $\psi_2(x)$ popisuje počáteční okamžitou rychlost bodů struny, tj. v čase $t = 0$.

Funkce $\varphi_1(t)$ a $\varphi_2(t)$ popisují kmitání konců struny. Funkce $f(x, t)$ má význam velikosti vnější síly působící v čase t v bodě x . Obecněji může jít o jakékoli vlnění v akustice, hydrodynamice, elektromagnetismu apod.

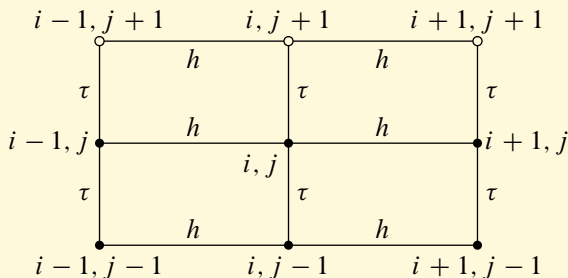
Řešení nemůžeme hledat na nekonečném časovém intervalu, proto se omezíme na interval $t \in \langle 0, T \rangle$, kde $T > 0$. Volíme obdélníkovou síť s kroky $h > 0$, $\tau > 0$, kde $h = l/n$, $\tau = T/m$, $m, n \in \mathbb{N}$. Tedy interval $\langle 0, l \rangle$ rozdělíme na n stejných dílků a interval $\langle 0, T \rangle$ na m stejných dílků — viz obr. 7.12. Označíme $u(x_i, t_j) = u_{ij}$, $i = 0, \dots, n$, $j = 0, \dots, m$.

Hodnoty v hraničních uzlech známe: $u_{i0} = \psi_1(x_i)$, $i = 0, \dots, n$, a $u_{0j} = \varphi_1(t_j)$, $u_{nj} = \varphi_2(t_j)$, $j = 0, \dots, m$. Na obr. 7.12 jsou tyto uzly vyznačeny plnými kroužky. Níže ještě určíme z druhé počáteční podmínky hodnoty v uzlech první vrstvy u_{i1} , $i = 1, \dots, n - 1$ (jsou vyznačeny plnými trojúhelníčky). Ve zbývajících uzlech (jsou vyznačeny prázdnými kroužky) musíme hodnoty řešení určit.



Obr. 7.12: Vlnová rovnice — volba sítě

Pro sestavení síťových rovnic použijeme následující třívrstvou devítibodovou šablonu:



V rovnici (7.18) nahradíme pomocí vztahu (7.3) derivace. Nahrazení u_{tt} bude přímočaré:

$$u_{tt}(x_i, t_j) = \frac{u_{i,j-1} - 2u_{ij} + u_{i,j+1}}{\tau^2} + O(\tau^2).$$

V případě u_{xx} budeme postupovat trochu složitěji. Platí

$$u_{xx}(x_i, t_{j-1}) = \frac{u_{i-1,j-1} - 2u_{i,j-1} + u_{i+1,j-1}}{h^2} + O(h^2),$$

$$u_{xx}(x_i, t_j) = \frac{u_{i-1,j} - 2u_{ij} + u_{i+1,j}}{h^2} + O(h^2),$$

$$u_{xx}(x_i, t_{j+1}) = \frac{u_{i-1,j+1} - 2u_{i,j+1} + u_{i+1,j+1}}{h^2} + O(h^2).$$

Nyní zvolíme reálné číslo σ a z posledních tří vztahů vytvoříme vážený průměr s vahami σ , $1 - 2\sigma$ a σ , kterým aproximujeme druhou derivaci u_{xx} . Konečně označíme $f_{ij} = f(x_i, t_j)$. Po zanedbání zbytků dostaneme pro přibližné hodnoty řešení v_{ij} vztah

$$\begin{aligned} & \frac{v_{i,j-1} - 2v_{ij} + v_{i,j+1}}{\tau^2} = \\ & = a^2 \left[\sigma \frac{v_{i-1,j+1} - 2v_{i,j+1} + v_{i+1,j+1}}{h^2} + (1 - 2\sigma) \frac{v_{i-1,j} - 2v_{ij} + v_{i+1,j}}{h^2} + \right. \\ & \quad \left. + \sigma \frac{v_{i-1,j-1} - 2v_{i,j-1} + v_{i+1,j-1}}{h^2} \right] + f_{ij}, \end{aligned}$$

z něhož po úpravě obdržíme tzv. *diferenční schéma s vahou*

$$\begin{aligned} v_{i,j+1} - \sigma \frac{a^2 \tau^2}{h^2} (v_{i-1,j+1} - 2v_{i,j+1} + v_{i+1,j+1}) &= \\ &= 2v_{ij} - v_{i,j-1} + (1 - 2\sigma) \frac{a^2 \tau^2}{h^2} (v_{i-1,j} - 2v_{ij} + v_{i+1,j}) + \\ &+ \sigma \frac{a^2 \tau^2}{h^2} (v_{i-1,j-1} - 2v_{i,j-1} + v_{i+1,j-1}) + \tau^2 f_{ij}. \end{aligned} \quad (7.21)$$

Tato lineární síťová rovnice váže přibližné hodnoty řešení v devíti bodech výše uvedené šablony.

Nyní se vrátíme k druhé počáteční podmínce $u_t(x, 0) = \psi_2(x)$, $0 \leq x \leq l$, a k určení přibližných hodnot řešení v uzlech první vrstvy, tj. v_{i1} , $i = 1, \dots, n-1$. Na obr. 7.12 jsou tyto uzly označeny trojúhelníčky. Ukážeme si dva způsoby. Budeme předpokládat, že potřebné funkce jsou dostatečně hladké.

Použijeme Taylorův vzorec druhého řádu na funkci $u(x_i, t)$ (proměnná je t , takže

derivace budou vlastně parciální derivace podle t) se středem 0. Pro $t = \tau$ dostaneme, že

$$u(x_i, \tau) = u(x_i, 0) + \tau u_t(x_i, 0) + \frac{\tau^2}{2} u_{tt}(x_i, 0) + O(\tau^3).$$

Protože je splněna vlnová rovnice (7.18), platí $u_{tt}(x_i, 0) = a^2 u_{xx}(x_i, 0) + f(x_i, 0)$. Po dosazení a úpravě vyjde:

$$\frac{u(x_i, \tau) - u(x_i, 0)}{\tau} = u_t(x_i, 0) + \frac{\tau a^2}{2} u_{xx}(x_i, 0) + \frac{\tau}{2} f(x_i, 0) + O(\tau^2). \quad (7.22)$$

Při prvním způsobu z počátečních podmínek obdržíme, že platí $u_t(x, 0) = \psi_2(x)$ a $u(x, 0) = \psi_1(x)$, tudíž $u_{xx}(x, 0) = \psi_1''(x)$. Po dosazení do předchozího vztahu, zanedbání zbytku a zjednodušení označení vyjde pro přibližné hodnoty řešení v nulté a první vrstvě lineární rovnice

$$\frac{v_{i1} - v_{i0}}{\tau} = \psi_2(x_i) + \frac{\tau a^2}{2} \psi_1''(x_i) + \frac{\tau}{2} f_{i0}. \quad (7.23)$$

Při druhém způsobu nahradíme v rovnosti (7.22) druhou derivaci $u_{xx}(x_i, 0)$ podle vztahu (7.3), tj.

$$u_{xx}(x_i, 0) = \frac{u_{i-1,0} - 2u_{i0} + u_{i+1,0}}{h^2} + O(h^2).$$

Po dosazení, zanedbání zbytku a zjednodušení označení vyjde pro přibližné hodnoty řešení v nulté a první vrstvě lineární rovnice

$$\frac{v_{i1} - v_{i0}}{\tau} = \psi_2(x_i) + \frac{\tau a^2}{2h^2} (v_{i-1,0} - 2v_{i0} + v_{i+1,0}) + \frac{\tau}{2} f_{i0}. \quad (7.24)$$

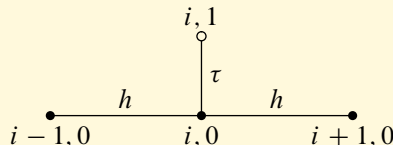
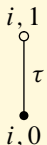
Protože hodnoty v_{00}, \dots, v_{n0} nulté vrstvy přesně známe (musí být splněna první počáteční podmínka), můžeme pomocí vztahu (7.23) resp. (7.24) snadno určit hodnoty v první vrstvě. Po úpravě dostaneme

$$v_{i1} = v_{i0} + \tau \psi_2(x_i) + \frac{\tau^2 a^2}{2} \psi_1''(x_i) + \frac{\tau^2}{2} f_{i0} \quad (7.25)$$

resp.

$$v_{i1} = v_{i0} + \tau \psi_2(x_i) + \frac{\tau^2 a^2}{2h^2} (v_{i-1,0} - 2v_{i0} + v_{i+1,0}) + \frac{\tau^2}{2} f_{i0}. \quad (7.26)$$

Postupně volíme $i = 1, \dots, n - 1$. Protože všechny hodnoty na pravých stranách obou předchozích rovností jsou známy, získáme hodnoty $v_{11}, \dots, v_{n-1,1}$ (hodnoty v_{01} a v_{n1} jsou přesně dány okrajovými podmínkami). Následující obrázek znázorňuje, které hodnoty nulté vrstvy (jsou označeny plnými kroužky) jsou použity pro určení v_{i1} pomocí vzorce (7.25) (levé schéma, jedna hodnota) resp. (7.26) (pravé schéma, obecně tři hodnoty).



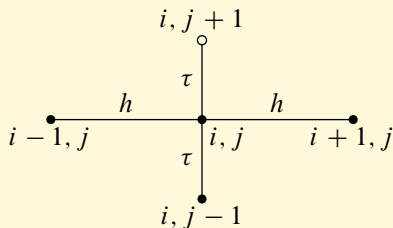
Dále rozlišíme dva případy.

Explicitní schéma ($\sigma = 0$)

Diferenční schéma (7.21) bude mít následující podobu:

$$v_{i,j+1} = 2v_{ij} - v_{i,j-1} + \frac{a^2 \tau^2}{h^2} (v_{i-1,j} - 2v_{ij} + v_{i+1,j}) + \tau^2 f_{ij}. \quad (7.27)$$

Levá strana obsahuje hodnotu pouze v jednom bodě ($j + 1$)-ní vrstvy, takže šablona se zjednoduší:



Protože známe hodnoty řešení v_{ij} v nulté a první vrstvě, tj. pro $j = 0$ a $j = 1$, a na levé a pravé straně obdélníku $\langle 0, l \rangle \times \langle 0, T \rangle$ — na obr. 7.12 jsou tyto uzly vyznačeny plnými kroužky a trojúhelníčky, umožňuje nám vzorec (7.27) počítat řešení po vrstvách. Stačí umístit základnu šablony na nultou vrstvu, tj. ve vzorci (7.27) zvolit $j = 1$, a např. odleva doprava postupně určit hodnoty neznámých $v_{12}, \dots, v_{n-1,2}$. Pak šablonu zvedneme o jednu vrstvu, tj. zvolíme $j = 2$, a obdobně určíme hodnoty neznámých $v_{13}, \dots, v_{n-1,3}$. A tak dál. Nevzniká tedy rozsáhlá soustava lineárních rovnic, kterou je nutné řešit jako celek. Proto se toto schéma nazývá *explicitní*.

Implicitní schéma ($\sigma \neq 0$)

Je-li $\sigma \neq 0$, nazývá se schéma (7.21) *implicitní* a situace je jiná. Levá strana vzorce (7.21) obsahuje s nenulovými koeficienty přibližné hodnoty řešení ve třech sousedních bodech $(j + 1)$ -první vrstvy. Pro $j = 1$ můžeme dosadit do pravé strany známé hodnoty v nulté a první vrstvě. Posouváme-li šablonu postupně zleva doprava, tj. volíme-li $i = 1$ až $n - 1$, dostaneme soustavu $n - 1$ lineárních rovnic pro neznámé $v_{12}, \dots, v_{n-1,2}$ v druhé vrstvě. Přitom první rovnice obsahuje pouze neznámé v_{12} a v_{22} (hodnotu v_{02} známe přesně), druhá rovnice obsahuje neznámé v_{12}, v_{22} a v_{32} atd. Předposlední rovnice obsahuje neznámé $v_{n-3,2}, v_{n-2,2}$ a $v_{n-1,2}$ a poslední pouze neznámé $v_{n-2,2}$ a $v_{n-1,2}$ (hodnotu v_{n2} opět známe přesně). Matice soustavy je tudíž třídiagonální a na řešení takových soustav existují velmi účinné metody.

Podobně postupujeme dále. Ve vzorci (7.21) zvolíme $j = 2$ a sestavíme stejným způsobem soustavu lineárních rovnic pro neznámé $v_{13}, \dots, v_{n-1,3}$. A tak dál. Na rozdíl od eliptického případu (Laplaceova a Poissonova rovnice) nemusíme podobně jako v parabolickém případě (rovnice vedení tepla) řešit jednu obrovskou soustavu pro všechny neznámé v_{ij} jako celek, dojde k rozpadnutí na samostatné soustavy pro každou vrstvu zvlášť.

Lze dokázat (viz [47, str. 260 a 315]), že pro dostatečně hladké funkce $f(x, t)$, $\psi_1(x)$, $\psi_2(x)$, $\varphi_1(t)$ a $\varphi_2(t)$ diferenční schéma (7.21) a (7.23) resp. (7.24) *aproximuje pro libovolné* σ úlohu (7.18) – (7.20) s řádem dva, tj. s přesností $O(h^2 + \tau^2)$.

Dále lze ukázat (viz [47, str. 313]), že pokud je $\sigma \geq 0$ a je splněna podmínka

$$\sigma \geq \frac{1 + \varepsilon}{4} - \frac{h^2}{4a^2\tau^2}, \quad (7.28)$$

kde $\varepsilon > 0$ je nějaké číslo, je schéma (7.21) a (7.23) *stabilní* (v jisté normě).

Tedy v případě *explicitního schématu* (7.27) a (7.23) je stabilita zajištěna podmínkou

$$\frac{a^2\tau^2}{h^2} \leq \frac{1}{1 + \varepsilon} \quad (7.29)$$

(říkáme, že toto schéma je *neabsolutně stabilní*). Tedy tato podmínka požaduje, aby podíl $a\tau/h$ byl menší než nějaké kladné číslo, které je menší než jedna. Schéma je pak konvergentní s řádem dvě (chyba je $O(\tau^2 + h^2)$).

Pro $\sigma > 1/4$ je podmínka (7.28) splněna pro libovolné kroky h a τ (schéma je *absolutně stabilní*), takže takové implicitní schéma (7.21) a (7.23) je konvergentní s řádem dvě (chyba je $O(\tau^2 + h^2)$).

Rovněž je možné ukázat (viz [47, str. 317]), že pokud je splněna podmínka

$$\sigma \geq \frac{1}{4} - \frac{h^2}{4a^2\tau^2}, \quad (7.30)$$

je schéma (7.21) a (7.24) *stabilní* (v jisté tzv. polonormě).

Tedy v případě *explicitního schématu* (7.27) a (7.24) je stabilita zajištěna podmínkou

$$\frac{a\tau}{h} \leq 1 \quad (7.31)$$

(říkáme, že toto schéma je neabsolutně stabilní). Schéma je pak konvergentní s řádem dvě (chyba je $O(\tau^2 + h^2)$). Podmínka (7.31) je známa jako *Courantova*¹.

Pro $\sigma \geq 1/4$ je podmínka (7.30) splněna pro libovolné kroky h a τ (schéma je absolutně stabilní), takže takové implicitní schéma (7.21) a (7.24) je konvergentní s řádem dvě (chyba je $O(\tau^2 + h^2)$).

¹**Richard Courant** (1888–1972) (čti kurant) — významný německý matematik. V r. 1922 založil na univerzitě v Göttingenu Matematický institut. Roku 1933 byl vyhnán pro neárijský původ a odešel do USA. Na newyorské univerzitě vytvořil výzkumné centrum pro aplikovanou matematiku. Byl spoluzakladatelem a prvním ředitelem Institutu matematických věd, který se stal jedním z nejprestižnějších světových matematických pracovišť; od r. 1964 nese jeho jméno ([Courant Institute of Mathematical Sciences](#)).

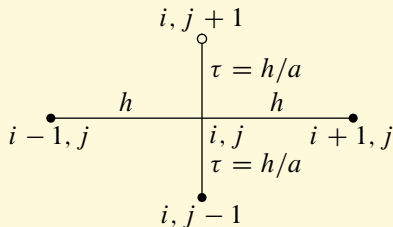
Explicitní schéma ($\sigma = 0$) se speciální volbou kroků

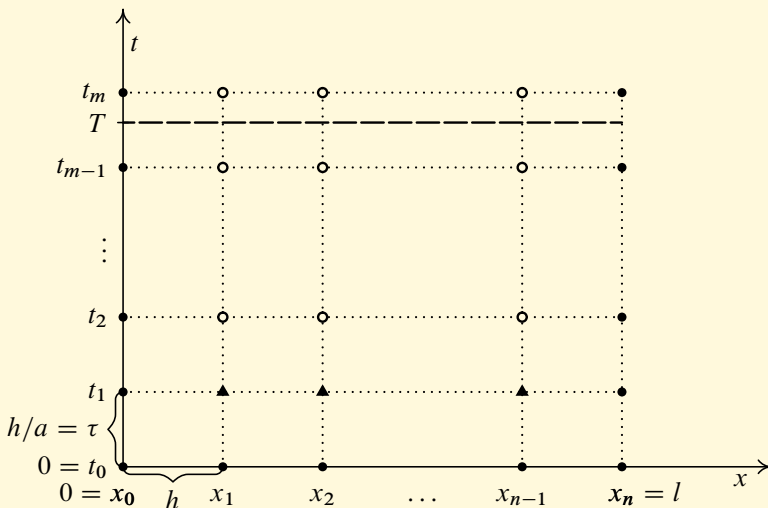
Podmínka (7.31) připouští volbu kroků h a τ takovou, že $\tau = h/a$. Podíváme se, jak se příslušné vzorce zjednoduší. Protože délka T časového intervalu obecně nebude při takové volbě délky kroku τ jejím celočíselným násobkem, zvolíme tedy počet dílků m na časové ose tak, aby $t_m = m\tau \geq T$, tj. abychom se dostali za čas T — viz obr. 7.13.

Dosadíme-li do vztahu (7.27) rovnost $a\tau = h$, dostaneme po úpravě, že

$$v_{i,j+1} = v_{i-1,j} - v_{i,j-1} + v_{i+1,j} + \tau^2 f_{ij}. \quad (7.32)$$

Šablona se tedy zjednoduší, síťová rovnice neobsahuje neznámou v_{ij} :



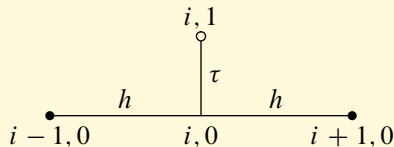


Obr. 7.13: Vlnová rovnice — speciální volba sítě

Podobně po dosazení do (7.26) vyjde po úpravě

$$v_{i1} = \frac{v_{i-1,0} + v_{i+1,0}}{2} + \tau \psi_2(x_i) + \frac{\tau^2}{2} f_{i0}. \quad (7.33)$$

Tento vztah použijeme k určení hodnot $v_{11}, \dots, v_{n-1,1}$ ze známých hodnot v_{00}, \dots, v_{n0} nulté vrstvy. Všimněte si, že vztah rovněž neobsahuje hodnotu v_{i0} . Následující obrázek znázorňuje, které dvě hodnoty nulté vrstvy jsou použity pro určení v_{i1} pomocí vzorce (7.33) (jsou označeny plnými kroužky).



Explicitní schéma (7.32) a (7.33) je (podmíněně) stabilní, a tudíž konvergentní s řádem dvě (chyba je $O(\tau^2 + h^2) = O(h^2/a^2 + h^2) = O(h^2)$).

Ukázka 7.3 Uvažujme vlnovou rovnici

$$u_{tt} = 3u_{xx}, \quad 0 \leq x \leq 3, \quad 0 \leq t < \infty,$$

s počátečními podmínkami

$$u(x, 0) = \frac{1}{6} (6x - x^2 - 8)^2 (x - 1)x, \quad u_t(x, 0) = 0, \quad 0 \leq x \leq 3,$$

a okrajovými podmínkami

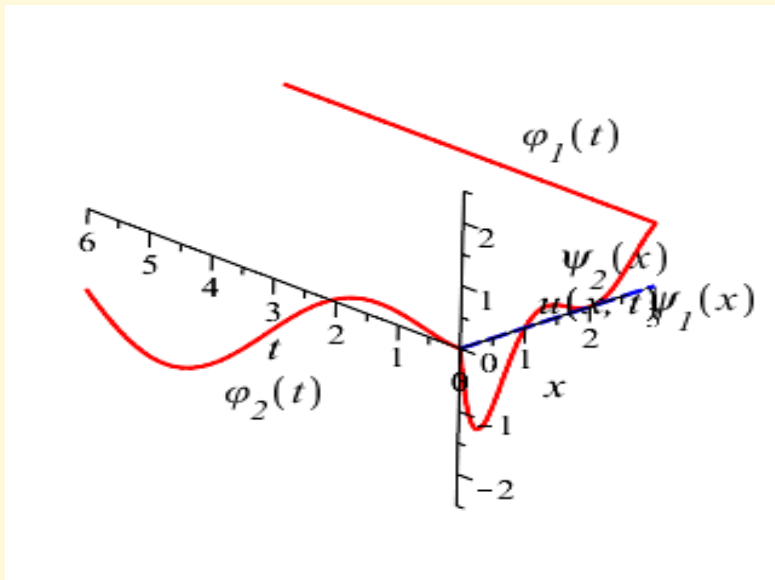
$$u(0, t) = \sin t + e^{-t} - 1, \quad u(3, t) = 1, \quad 0 \leq t < \infty.$$

Při označení z (7.18) platí, že $a = \sqrt{3}$ a $l = 3$. Dále je $\psi_1(x) = \frac{1}{6} (6x - x^2 - 8)^2 (x - 1)x$, $\psi_2(x) = 0$, $\varphi_1(t) = \sin t + e^{-t} - 1$, $\varphi_2(t) = 1$ a $f(x, t) = 0$. Rovnice může popisovat např. kmitání struny.

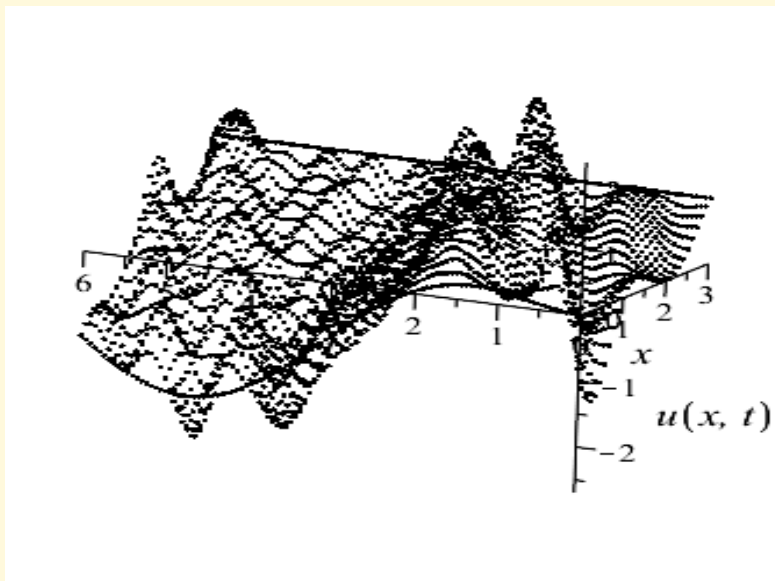
Použijeme implicitní schéma (7.21) pro $\sigma = 1$ a (7.26). Zvolili jsme $T = 6$, tedy řešení bylo nalezeno na obdélníku $\langle 0, 3 \rangle \times \langle 0, 6 \rangle$. Dále jsme vybrali $n = 30$, $m = 104$ takže $h = 3/30 = 0,1$ a $\tau = 6/104 \doteq 0,058$. Použité schéma je vždy konvergentní.

Počet bodů sítě je $31 \cdot 105 = 3\,255$. Odečteme-li hraniční body, v nichž známe přesné hodnoty řešení, zůstane 3 016 neznámých. Nejprve pomocí vztahu (7.26) vypočítáme 29 neznámých $v_{1,1}, \dots, v_{29,1}$ v první vrstvě. Pak pomocí vztahu (7.21) vypočítáme zbývající neznámé. Vzniklou soustavu síťových rovnic nemusíme řešit jako celek, rozpadne se na 103 menších soustav odpovídajících neznámým v jednotlivých vrstvách, které mají po 29 neznámých. Neznámé udávají přibližné hodnoty v_{ij} řešení v uzlových bodech.

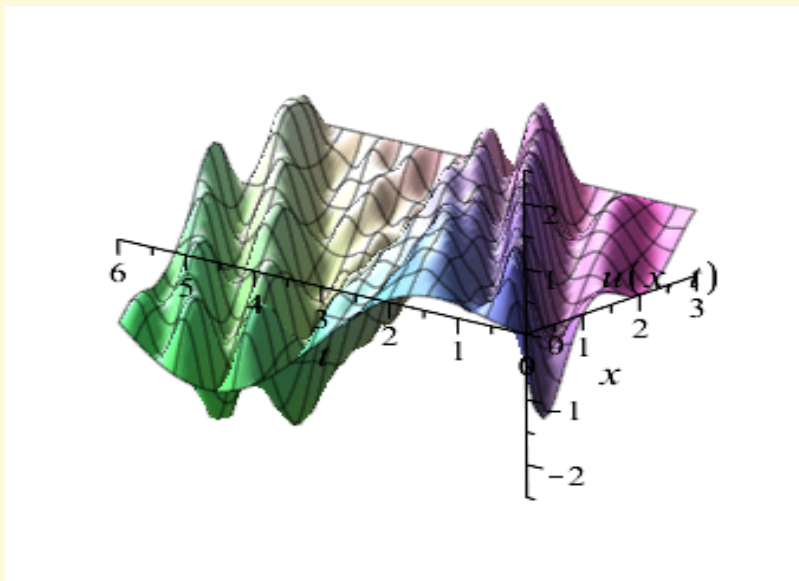
Na obr. 7.14 jsou zobrazeny okrajové a počáteční podmínky a na obr. 7.15 jsou zobrazené přibližné hodnoty řešení $u(x, t)$ této úlohy v uzlových bodech. Na obr. 7.16 je těmito body proložena plocha dávající představu o řešení na celém obdélníku. Na obr. 7.17 je animace znázorňující kmitání struny v časovém intervalu $\langle 0, 6 \rangle$.



Obr. 7.14: Řešení vlnové rovnice — okrajové a počáteční podmínky



Obr. 7.15: Řešení vlnové rovnice — hodnoty řešení v uzlových bodech



Obr. 7.16: Graf řešení vlnové rovnice

Obr. 7.17: Animace řešení vlnové rovnice (kmity struny)

Pojmy k zapamatování

- lineární parciální diferenciální rovnice druhého řádu
- princip metody sítí
- síť, hraniční a vnitřní body sítě
- šablona, diferenční schéma, síťové rovnice
- korektnost a stabilita diferenčního schématu
- diferenční schéma aproximující úlohu
- konvergentní diferenční schéma
- náhrada derivací diferencemi
- Dirichletova úloha pro Poissonovu rovnici
- šablona pro Poissonovu rovnici
- přenos okrajových podmínek
- konvergence diferenčního schématu pro Poissonovu rovnici
- smíšená úloha pro rovnici vedení tepla

- šablona pro rovnici vedení tepla, explicitní a implicitní schémata
- stabilita a konvergence schémat pro rovnici vedení tepla
- smíšená úloha pro vlnovou rovnici
- šablona pro vlnovou rovnici, explicitní a implicitní schémata
- stabilita a konvergence schémat pro vlnovou rovnici

Kontrolní otázky

1. Co rozumíme sítí a s jakými typy sítí se můžeme setkat?
2. Co je výsledkem řešení parciálních diferenciálních rovnic metodou sítí?
3. Které uzly se nazývají hraniční a které vnitřní?
4. Co je to šablona?
5. Popište postup řešení PDR metodou sítí.
6. Co jsou to síťové rovnice?
7. Co je to diferenční schéma?
8. Kdy je diferenční schéma korektní?
9. Co znamená, že diferenční schéma aproximuje nějakou parciální diferenciální rovnici s okrajovými a počátečními podmínkami?
10. Co znamená, že diferenční schéma je konvergentní? Uveďte postačující podmínku konvergence diferenčního schématu.
11. Čím nahrazujeme první a druhé parciální derivace při řešení parciálních diferenciálních rovnic metodou sítí?
12. Napište, jak vypadá Poissonova rovnice.

13. Napište, jak vypadá Laplaceova rovnice.
14. Jakou úlohu označujeme Dirichletova?
15. Jak vypadá šablona pro řešení Poissonovy rovnice metodou sítí?
16. Které podmínky označujeme jako počáteční a které jako okrajové?
17. Napište, jak vypadá rovnice smíšené úlohy pro vedení tepla.
18. Jak vypadá síť pro rovnici vedení tepla?
19. Nakreslete, jak vypadá šablona pro řešení smíšené úlohy pro vedení tepla metodou sítí v případě implicitního, explicitního a ryze implicitního schématu.
20. Co je Crankovo-Nicolsonové schéma a čím se odlišuje od ostatních schémat?
21. Napište podmínku pro stabilitu řešení úlohy pro vedení tepla v případě implicitního schématu.
22. Napište podmínku pro stabilitu řešení úlohy pro vedení tepla v případě explicitního schématu.
23. Jak vypadá postup řešení smíšené úlohy pro vedení tepla metodou sítí? Čím se liší použití explicitního a implicitního schématu?
24. Napište, jak vypadá rovnice smíšené úlohy pro vlnovou rovnici.
25. Jak vypadá síť pro řešení vlnové rovnice?

26. Nakreslete, jak vypadá šablona pro řešení smíšené úlohy pro vlnovou rovnici metodou sítí v případě implicitního a explicitního schématu.
27. Co je explicitní schéma pro vlnovou rovnici se speciální volbou kroků?
28. Jak vypadá postup řešení smíšené úlohy pro vlnovou rovnici metodou sítí? Čím se liší použití explicitního a implicitního schématu?
29. Napište podmínku pro stabilitu řešení vlnové rovnice v případě implicitního schématu.
30. Napište podmínku pro stabilitu řešení vlnové rovnice v případě explicitního schématu.
31. Jaký je rozdíl při použití explicitních a implicitních diferenčních schémat pro rovnici vedení tepla resp. vlnovou rovnici? Srovnajte situaci s řešením Poissonovy rovnice.

Testy ke kapitole 7

Vyberte správnou odpověď (právě jedna je správná). Za chybnou odpověď se neodečítají body. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (1b.) Vyberte, která odpověď nejlépe charakterizuje, co je síť.

Je to množina křivek, které spojují uzlové body; dva uzly mohou být spojeny nejvýše jednou takovou křivkou.

Je to konečná množina bodů, která pokrývá definiční obor uvažované parciální diferenciální rovnice.

Je to konečná množina trojúhelníků, jejichž sjednocení pokrývá definiční obor uvažované parciální diferenciální rovnice.

Je to konečná množina obdélníků, jejichž sjednocení pokrývá definiční obor uvažované parciální diferenciální rovnice.

2. (1b.) Jaký je princip numerického řešení metodou sítí?

Najdeme přesné hodnoty řešení uvažované parciální diferenciální rovnice v uzlech sítě.

Najdeme analytické řešení uvažované parciální diferenciální rovnice.

Najdeme přibližné analytické řešení uvažované parciální diferenciální rovnice platné v celém definičním oboru. Jeho přesnost je však uspokojující pouze v uzlových bodech.

Najdeme přibližné hodnoty řešení uvažované parciální diferenciální rovnice v uzlech sítě.

3. (1b.) Parciální diferenciální rovnice $u_{xx} + u_{yy} = 0$ se nazývá

vlnová rovnice.

Müllerova.

Mählerova.

Møellerova.

Poissonova.

Laplaceova.

Fourierova.

Dirichletova.

rovnice vedení tepla.

4. (1b.) Co to znamená, že diferenční schéma $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$ je korektní?

Pro libovolné malé $\|\mathbf{h}\|$ je dobře podmíněné.

Je podmíněně konvergentní.

Pro dostatečně malé $\|\mathbf{h}\|$ má pro libovolnou pravou stranu $B_{\mathbf{h}}$ jediné řešení a je stabilní.

Pro dostatečně malé $\|\mathbf{h}\|$ je matice $A_{\mathbf{h}}$ regulární.

5. (1b.) Co to znamená, že diferenční schéma $A_{\mathbf{h}} V_{\mathbf{h}} = B_{\mathbf{h}}$ je konvergentní?

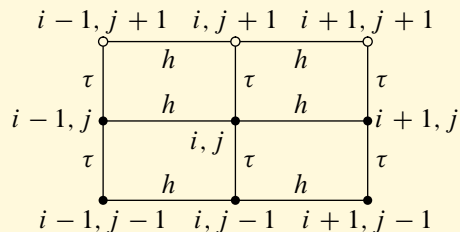
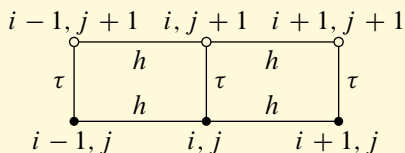
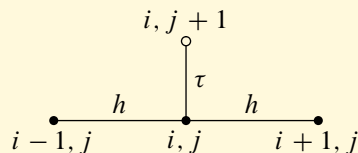
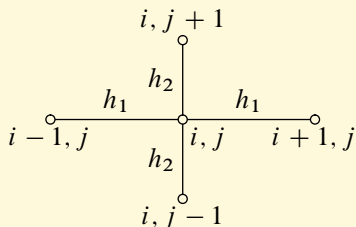
Musí platit, že $\|V_{\mathbf{h}} - U_{\mathbf{h}}\| \rightarrow 0$ pro $\|\mathbf{h}\| \rightarrow 0$, kde $U_{\mathbf{h}}$ je sloupec přesných hodnot řešení v bodech sítě.

Musí platit, že $\|B_{\mathbf{h}} - A_{\mathbf{h}} U_{\mathbf{h}}\| \rightarrow 0$ pro $\|\mathbf{h}\| \rightarrow 0$, kde $U_{\mathbf{h}}$ je sloupec přesných hodnot řešení v bodech sítě.

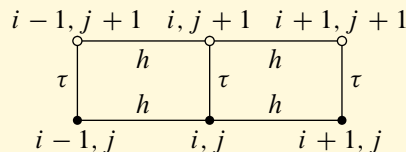
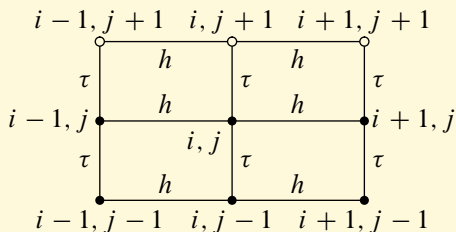
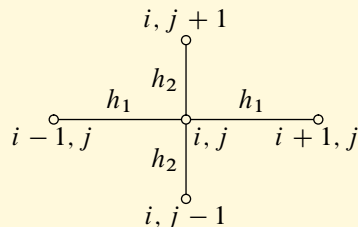
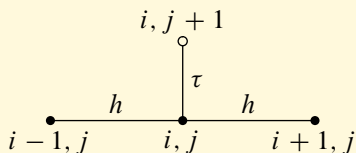
Musí existovat kladná konstanta M taková, že $\|A_{\mathbf{h}}^{-1}\| \leq M$ pro libovolné dostatečně malé $\|\mathbf{h}\|$.

Sloupec hodnot $V_{\mathbf{h}}$ v bodech sítě konverguje k hodnotám přibližného řešení.

6. (1b.) Která šablona odpovídá obecnému diferenčnímu schématu pro smíšenou úlohu pro rovnici vedení tepla?



7. (1b.) Která šablona odpovídá obecnému diferenčnímu schématu pro smíšenou úlohu pro vlnovou rovnici?



8. (1b.) Které tvrzení o soustavě síťových rovnic získané při řešení smíšené úlohy pro rovnici vedení tepla pomocí Crankova-Nicolsonové schématu je správné?

Soustavu není možné řešit přímými metodami.

Soustavu se rozpadne na menší soustavy pro jednotlivé vrstvy.

Soustavu je nutné řešit jako celek, je ireducibilní.

Řešení soustavy v podstatě odpadne, hodnoty jednotlivých neznámých počítáme postupně po vrstvách.

9. (1b.) Které tvrzení o soustavě síťových rovnic získané při řešení smíšené úlohy pro vlnovou rovnici pomocí explicitního schématu je správné?

Soustavu se rozpadne na menší soustavy pro jednotlivé vrstvy.

Soustavu není možné řešit iteračními metodami.

Řešení soustavy v podstatě odpadne, hodnoty jednotlivých neznámých počítáme postupně po vrstvách.

Soustavu je nutné řešit jako celek, je ireducibilní.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (1b.) Uzly sítě dělíme na

externí a interní.

otevřené a uzavřené.

hraniční a vnitřní.

uzly se sudou a lichou násobností.

2. (1b.) Co jsou to síťové rovnice?

Jsou to lineární algebraické rovnice vzniklé diskretizací uvažované parciální diferenciální rovnice v uzlech sítě.

Jsou to nelineární algebraické rovnice vzniklé diskretizací uvažované parciální diferenciální rovnice v uzlech sítě.

Jsou to obyčejné lineární diferenciální rovnice získané zafixováním jedné nezávisle proměnné.

Jsou to obyčejné nelineární diferenciální rovnice získané zafixováním jedné nezávisle proměnné.

3. (1b.) Co to znamená, že diferenční schéma $A_h V_h = B_h$ je stabilní?

Pro dostatečně malé $\|h\|$ je matice A_h regulární.

Pro libovolné malé $\|h\|$ je dobře podmíněné.

Musí existovat kladná konstanta M taková, že $\|A_h^{-1}\| \leq M$ pro libovolné dostatečně malé $\|h\|$.

Je podmíněně konvergentní.

4. (1b.) Jaké podmínky zaručují, že diferenční schéma $A_h V_h = B_h$ je konvergentní?

Když je diferenční schéma korektní a konzistentní.

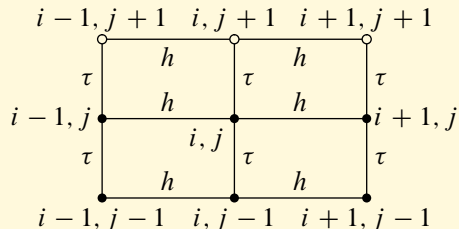
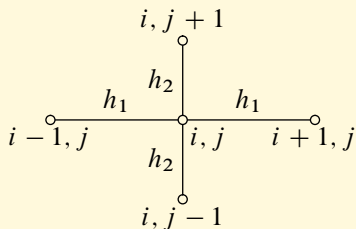
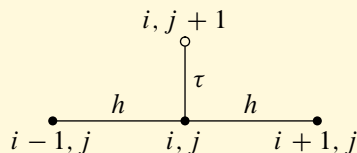
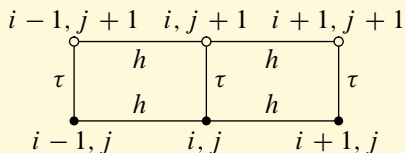
Když je diferenční schéma stabilní a aproximuje uvažovanou parciální diferenciální rovnici.

Když je diferenční schéma korektní a stabilní.

Když je diferenční schéma dobře podmíněné.

5. (1b.) Parciální diferenciální rovnice $u_t = a^2 u_{xx} + f(x, t)$ se nazývá
- | | | |
|---------------|-----------------------|-----------------|
| Dirichletova. | Poissonova. | vlnová rovnice. |
| Müllerova. | rovnice vedení tepla. | Fourierova. |
| Mählerova. | Møellerova. | Laplaceova. |
6. (1b.) Které tvrzení o soustavě síťových rovnic získané při řešení Dirichletovy úlohy pro Poissonovu rovnici na obdélníku je správné?
- Soustavu je nutné řešit jako celek, je ireducibilní.
 - Soustavu je možné řešit po vrstvách.
 - Soustavu je možné řešit po sloupcích.
 - Soustavu není možné řešit iteračními metodami.

7. (1b.) Která šablona odpovídá explicitnímu diferenčnímu schématu pro smíšenou úlohu pro rovnici vedení tepla?



8. (1b.) Které tvrzení o soustavě síťových rovnic získané při řešení smíšené úlohy pro vlnovou rovnici pomocí diferenčního schématu s nenulovou vahou je správné?

Řešení soustavy v podstatě odpadne, hodnoty jednotlivých neznámých počítáme postupně po vrstvách.

Soustavu není možné řešit přímými metodami.

Soustavu se rozpadne na menší soustavy pro jednotlivé vrstvy.

Soustavu je nutné řešit jako celek, je ireducibilní.

9. (1b.) Co znamená, že diferenční schéma je absolutně stabilní?

Schéma je stabilní, i když okrajové podmínky uvažované úlohy nahradíme jejich absolutními hodnotami.

Když aproximuje uvažovanou úlohu aspoň s řádem dvě.

Když podmínky zajišťující stabilitu automaticky zajišťují i aproximaci uvažované úlohy.

Když je stabilní při libovolné délce kroků.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (1b.) Vyberte, která odpověď nejlépe charakterizuje, co je šablona.

Je to postup, který se používá k řešení uvažované parciální diferenciální rovnice.

Je to obdélníkové schéma, které se používá k zadání počátečních a okrajových podmínek uvažované parciální diferenciální rovnice.

Je to skupina blízkých bodů sítě, na niž se provádí diskretizace uvažované parciální diferenciální rovnice.

Je to nepravidelné schéma, které se používá k přenosu počátečních nebo okrajových podmínek uvažované parciální diferenciální rovnice.

2. (1b.) Parciální diferenciální rovnice $u_{xx} + u_{yy} = f(x, y)$, $f(x, y) \neq 0$, se nazývá

Mählerova.

Møellerova.

Poissonova.

Fourierova.

Dirichletova.

rovnice vedení tepla.

vlnová rovnice.

Müllerova.

Laplaceova.

3. (1b.) Jaký je rozdíl mezi pojmy diferenční schéma a síťová rovnice?

Síťové rovnice jsou speciálním případem diferenčních schémat.

Diferenční schémata jsou speciálním případem síťových rovnic.

Je to totéž.

Tyto dva pojmy spolu nijak nesouvisí.

4. (1b.) Co to znamená, že diferenční schéma $A_h V_h = B_h$ aproximuje nějakou parciální diferenciální rovnici 2. řádu?

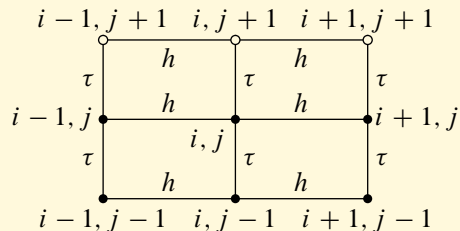
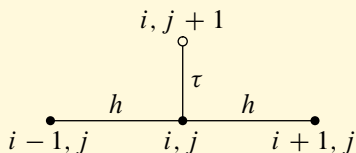
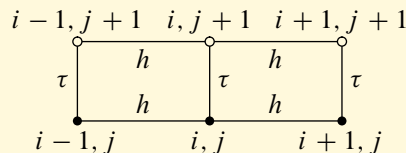
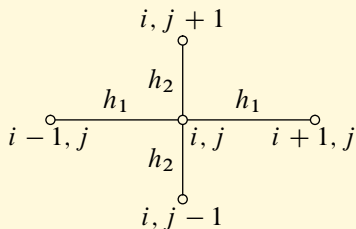
Musí platit, že $\|V_h - U_h\| \rightarrow 0$ pro $\|h\| \rightarrow 0$, kde U_h je sloupec přesných hodnot řešení v bodech sítě.

Musí existovat kladná konstanta M taková, že $\|A_h^{-1}\| \leq M$ pro libovolně dostatečně malé $\|h\|$.

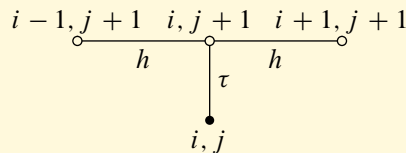
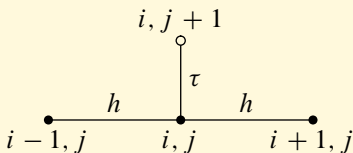
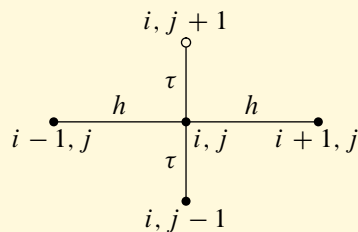
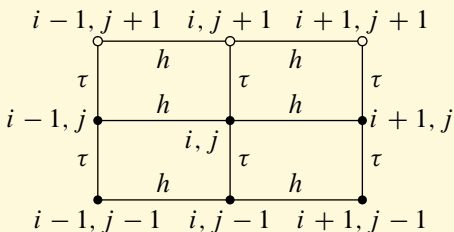
Musí být korektní a stabilní.

Musí platit, že $\|B_h - A_h U_h\| \rightarrow 0$ pro $\|h\| \rightarrow 0$, kde U_h je sloupec přesných hodnot řešení v bodech sítě.

5. (1b.) Která šablona se používá pro diskretizaci Dirichletovy úlohy pro Poissonovu rovnici?



6. (1b.) Která šablona odpovídá explicitnímu diferenčnímu schématu pro smíšenou úlohu pro vlnovou rovnici?



7. (1b.) Které tvrzení o soustavě síťových rovnic získané při řešení smíšené úlohy pro rovnici vedení tepla pomocí ryze implicitního schématu je správné?

Řešení soustavy v podstatě odpadne, hodnoty jednotlivých neznámých počítáme postupně po vrstvách.

Soustavu se rozpadne na menší soustavy pro jednotlivé vrstvy.

Soustavu je nutné řešit jako celek, je ireducibilní.

Soustavu není možné řešit přímými metodami.

8. (1b.) Které tvrzení o soustavě síťových rovnic získané při řešení smíšené úlohy pro rovnici vedení tepla pomocí explicitního schématu je správné?

Řešení soustavy v podstatě odpadne, hodnoty jednotlivých neznámých počítáme postupně po vrstvách.

Soustavu je nutné řešit jako celek, je ireducibilní.

Soustavu se rozpadne na menší soustavy pro jednotlivé vrstvy.

Soustavu není možné řešit přímými metodami.

9. (1b.) Je explicitní diferenční schéma pro smíšenou úlohu pro vlnovou rovnici absolutně stabilní?

Ano.

Ano, ale jen v případě homogenní rovnice.

Ne.

Bez znalosti délky kroků to nelze říci.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Souhrnné testy

Vyberte správné odpovědi (může být více správných odpovědí). Za chybný výběr se strhává bod. Test lze kdykoli tlačítky na konci ukončit a nechat si vypsát správné odpovědi.

Test 1

1. (3b.) Kvadraturní formule $Q(f) = \frac{1}{3}f(-1) + \frac{4}{3}f(0) + \frac{1}{3}f(1)$, nahrazující integrál $\int_{-1}^1 f(x) dx$, má stupeň přesnosti $N = 3$. Vyberte polynom, jehož integrál lze touto formulí určit s nulovou chybou.

Žádný takový polynom neexistuje.

$f(x) = 1 - x^3 + 3x.$ $f(x) = x^2 - 3x + 2.$

$f(x) = x^3 + 9x - 4.$ $f(x) = x^4 - 3x + 2.$ $f(x) = x^5 + x^3 - 1.$

2. (2b.) Máme nalézt kořeny rovnice $x \sin x + 2x - x^3 + 4 = 0$. Vyberte takové funkce $f_1(x)$ a $f_2(x)$, aby x -ové souřadnice průsečíků jejich grafů odpovídaly kořenům dané rovnice.

$$f_1(x) = x \sin x \text{ a } f_2(x) = -x^3 + 2x + 4.$$

$$f_1(x) = x \sin x + 2x \text{ a } f_2(x) = -x^3 + 4.$$

$$f_1(x) = x \sin x \text{ a } f_2(x) = x^3 - 2x - 4.$$

$$f_1(x) = -x \text{ a } f_2(x) = \frac{x \sin x - x^3 + 4}{2}.$$

3. (1b.) Je Crankovo-Nicolsonové diferenční schéma pro smíšenou úlohu pro rovnici vedení tepla absolutně stabilní?

Ano.

Ne.

Bez znalosti délky kroků to nelze říci.

Záleží to na okrajových podmínkách.

4. (1b.) Nechť se soustava $A\mathbf{x} = B$ dá řešit metodou LU rozkladu a nechť platí $A = LU$. Vyberte správné tvrzení.

Matice L je dolní trojúhelníková matice a U je horní trojúhelníková matice. Matice získáme z matice A diagonálním řezem, přičemž matice L má na diagonále samé jedničky.

Matice L je dolní trojúhelníková matice a U je horní trojúhelníková matice. Matice získáme z matice A rozdělením podle diagonály, přičemž prvky na diagonále přidáme do matice U .

Matice L je sestavena pomocí multiplikátorů při použití algoritmu GEM a U je výsledná horní trojúhelníková matice získaná příslušnými řádkovými úpravami.

5. (1b.) Definice oblasti absolutní stability numerické metody pro řešení ODR je založena na chování numerického řešení počáteční úlohy pro testovací rovnici

$$y' = \frac{\lambda}{y}, y(0) = 1.$$

$$y' = \lambda y, y(0) = 1.$$

$$y' = \lambda y^2, y(0) = 1.$$

$$y' = \lambda + y, y(0) = 1.$$

6. (2b.) Pro speciální Lagrangeovy polynomy

$$L_i(x) = \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}$$

platí:

$$\text{st}L_i(x) = n \text{ a } L_i(x_j) = 0 \text{ pro } i \neq j, \text{ kde } i, j = 1, \dots, n.$$

$$\text{st}L_i(x) = n \text{ a } L_i(x_i) = 0, L_i(x_j) = 1 \text{ pro } i \neq j.$$

$$\text{st}L_i(x) = n \text{ a } L_i(x_i) = 1, i = 1, \dots, n.$$

$$\text{st}L_i(x) = n - 1 \text{ a } L_i(x_i) = 1, L_i(x_j) = 0 \text{ pro } i \neq j.$$

7. (1b.) Maticová norma indukovaná nějakými vektorovými normami se nazývá
operátorová norma. součtová norma.

podmíněná norma.

Frobeniova norma.

8. (1b.) V každém z n uzlových bodů je předepsána funkční hodnota a hodnoty derivací až do řádu nejvýše m , $m \in \mathbb{N}$. Doplňte správné tvrzení: Pak Hermitův interpolační polynom

je těmito podmínkami jednoznačně určen.

není těmito podmínkami jednoznačně určen, je nutné přidat dvě okrajové podmínky.

není těmito podmínkami jednoznačně určen, je nutné přidat dvě okrajové podmínky a jeden uzel.

pro některé podmínky nemusí existovat.

9. (1b.) Vyberte předpis pro obecnou explicitní lineární k -krokovou metodu s konstantním krokem.

$$y_{i+1} = -a_1 y_i - \cdots - a_k y_{i+1-k} + h(b_0 f_{i+1} + \cdots + b_k f_{i+1-k}), b_0 \neq 0.$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}; f).$$

$$y_{i+1} = -a_1 y_i - \cdots - a_k y_{i+1-k} + h(b_1 f_i + \cdots + b_k f_{i+1-k}).$$

$$y_{i+1} = y_i + h_{i+1} \Phi(x_i, y_i, h_{i+1}, y_{i+1}; f).$$

10. (1b.) Globální diskretizační chyba metody po i krocích je definována jako rozdíl mezi přesnou hodnotou $y(x_i)$ a přibližnou hodnotou y_i . Nebereme-li v úvahu zaokrouhlovací chyby a zmenšíme-li krok h , pak

se tato chyba zmenší.

se tato chyba zvětší.

tato změna neovlivní velikost této chyby.

o její změně nelze nic říct, závisí na konkrétní diferenciální rovnici a metodě.

11. (3b.) Která numerická úloha je obecně špatně podmíněná?

Numerický výpočet první derivace.

Numerický výpočet určitého integrálu.

Numerický výpočet druhé derivace.

Každá numerická úloha.

Numerický výpočet derivace libovolného řádu.

Žádná numerická úloha.

12. (1b.) Vektorová norma $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$ se nazývá

Frobeniova.

maximální.

součtová.

Eukleidovská.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 2

1. (2b.) Která tvrzení o Rombergově kvadratuře jsou pravdivá?

Je založena na asymptotickém vzorci pro složené obdélníkové pravidlo.

Používá se, když požadujeme vysokou přesnost výsledku. Funkce musí být dostatečně hladká

Dává velmi přesné výsledky i pro nespojitě funkce.

Je založena na asymptotickém vzorci pro složené lichoběžníkové pravidlo.

2. (1b.) Součtová, Frobeniova a maximální maticová norma jsou analogiemi vektorových norem. součtových norem.

podmíněných norem.

operátorových norem.

3. (1b.) Čím se liší Jacobiova a Gaussova-Seidelova metoda?

Při určování i -té složky $(k + 1)$ -ní iterace $x_i^{(k+1)}$ používáme u Jacobiovy metody hodnoty $x_j^{(k)}$, $j \neq i$, kdežto u Gaussovy-Seidelovy metody hodnoty $x_j^{(k)}$ pro $j < i$ a $x_j^{(k+1)}$ pro $j > i$.

Při určování i -té složky $(k + 1)$ -ní iterace $x_i^{(k+1)}$ používáme u Jacobiovy metody hodnoty $x_j^{(k)}$, $j \neq i$, kdežto u Gaussovy-Seidelovy metody hodnoty $x_j^{(k+1)}$ pro $j < i$ a $x_j^{(k)}$ pro $j > i$.

Při určování i -té složky $(k + 1)$ -ní iterace $x_i^{(k+1)}$ používáme u Gaussovy-Seidelovy metody hodnoty $x_j^{(k)}$, $j \neq i$, kdežto u Jacobiovy metody hodnoty $x_j^{(k+1)}$ pro $j < i$ a $x_j^{(k)}$ pro $j > i$.

Při určování i -té složky $(k + 1)$ -ní iterace $x_i^{(k+1)}$ používáme u Gaussovy-Seidelovy metody hodnoty $x_j^{(k)}$, $j \neq i$, kdežto u Jacobiovy metody hodnoty $x_j^{(k)}$ pro $j < i$ a $x_j^{(k+1)}$ pro $j > i$.

4. (1b.) Číslo podmíněnosti matice A určíme ze vztahu

$$\kappa(A) = \|A\| \cdot \|A^T\|.$$

$$\kappa(A) = |A| \cdot |A^{-1}|.$$

$$\kappa(A) = |A| \cdot |A^T|.$$

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|.$$

5. (1b.) V případě, že funkci $f(x)$ nahradíme interpolačním polynomem $P(x)$ sestrojeným z jejích hodnot v uzlových bodech $x_i, i = 0, \dots, n$, potom platí:

Funkce $f(x)$ a polynom $P(x)$ mají v uzlových bodech $x_i, i = 0, \dots, n$, stejné funkční hodnoty.

Funkce $f(x)$ a polynom $P(x)$ mají v uzlových bodech $x_i, i = 0, \dots, n$, stejné funkční hodnoty a stejné hodnoty prvních n derivací.

Funkce $f(x)$ a polynom $P(x)$ nemusí mít v uzlových bodech $x_i, i = 0, \dots, n$, stejné funkční hodnoty.

Funkce $f(x)$ a polynom $P(x)$ nemají v uzlových bodech $x_i, i = 0, \dots, n$, stejné funkční hodnoty.

6. (3b.) Vyberte podmínky, které jsou ekvivalentní s tvrzením, že matice $A \in \mathbb{M}_n$ je regulární.

K matici A existuje matice inverzní A^{-1} . $A^T A = A A^T = E$, kde E je jednotková matice.

$|A| \neq 0$. $h(A) = n$.

7. (1b.) Funkci $f(x)$ máme aproximovat kvadratickým polynomem ve smyslu metody nejmenších čtverců. Vyberte správný předpis aproximující funkce $\varphi(x)$.

$\varphi(x) = a + bx^2$. $\varphi(x) = ax + bx^2$.

$\varphi(x) = a + bx + cx^2$. $\varphi(x) = a + bx$.

8. (1b.) Pomocí Eulerovy explicitní metody nalezněte přibližnou hodnotu y_1 přesné hodnoty $y(1,1)$ pro počáteční úlohu $y' = \frac{2y+x}{3x^2}$, $y(1) = 2$, s krokem $h = 0,1$.

$y_1 = \frac{6}{13}$. $y_1 = \frac{7}{3}$. $y_1 = \frac{9}{4}$. $y_1 = \frac{13}{6}$.

Úlohu nelze vyřešit, neznáme interval, na kterém řešení hledáme.

9. (3b.) Vyberte metody, která patří mezi lineární více krokové metody.

Adamsovy metody.

Metody Rungeho-Kutty.

Metody prediktor-korektor.

Metody zpětného derivování.

10. (1b.) Parciální diferenciální rovnice $u_{tt} = a^2 u_{xx} + f(x, t)$ se nazývá

Fourierova.

Dirichletova.

Poissonova.

rovnice vedení tepla.

Müllerova.

Mählerova.

vlnová rovnice.

Møellerova.

Laplaceova.

11. (2b.) Vyberte pravdivé tvrzení o úloze numerické integrace.

Chyba způsobená nepřesnostmi ve vstupních hodnotách je nepodstatná.

Je to obecně dobře podmíněná úloha.

Je to obecně špatně podmíněná úloha.

Rozhodující podíl na celkové chybě má chyba způsobená nepřesnostmi ve vstupních hodnotách.

12. (3b.) Vyberte pravdivé tvrzení o úloze numerické derivace.

Chyba způsobená nepřesnostmi ve vstupních hodnotách je nepřímo úměrná mocnině délky kroku.

Je to obecně dobře podmíněná úloha.

Je to špatně podmíněná úloha, jen když počítáme první derivace.

Je to obecně špatně podmíněná úloha.

Chyba numerického modelu je přímo úměrná mocnině délky kroku.

Je to dobře podmíněná úloha, když počítáme derivace sudého řádu.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Test 3

1. (2b.) Vypočtete Frobeniovu maticovou normu matice

$$A = \begin{pmatrix} 3 & 10 \\ -4 & -5 \end{pmatrix}.$$

$\sqrt{22}.$

$\sqrt{150}.$

10.

-4.

$5\sqrt{6}.$

2. (2b.) Hledáme-li kořeny rovnice $f(x) = 0$, pak hledáme x -ové souřadnice průsečíků grafu funkce $y = f(x)$ s osou x .

průsečíky funkce $y = f(x)$ s osou y .

nenulová čísla α taková, že $f(\alpha) = 0$.

reálná nebo komplexní čísla α , pro něž platí, že $f(\alpha) = 0$.

3. (1b.) Proč se v některých případech používá algoritmus GEM s částečným nebo úplným výběrem hlavních prvků?

Protože jsou tyto algoritmy méně náročné na dobu výpočtu než GEM bez výběru hlavního prvku.

Protože tyto algoritmy jsou obecně sofistikovanější než GEM bez výběru hlavního prvku a jsou výrazně rychlejší.

Protože při použití GEM bez výběru hlavního prvku není tento algoritmus obecně numericky stabilní.

Protože pro soustavy s velkým počtem rovnic jsou tyto algoritmy vhodnější.

4. (1b.) Vyberte nepravdivou odpověď: Kvadrurní formule, jejichž koeficienty A_i lze získat integrací interpolačního polynomu, který je určen body $[x_i, f(x_i)]$, $i = 0, \dots, n$, se nazývají

Gaussovy kvadrurní formule.

Lagrangeovy kvadrurní formule.

Newtonovy-Cotesovy kvadrurní formule.

5. (1b.) Pomocí Lagrangeova interpolačního polynomu

$$L(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1)$$

vypočtete hodnotu $L(2)$, jsou-li příslušné údaje dány následující tabulkou:

x_i	1,8	2,3
$f(x_i)$	2	3

$$0. \qquad 2. \qquad \frac{12}{5} \cdot \qquad \frac{13}{5} \cdot$$

6. (1b.) Definice D-stability lineární více krokové metody je založena na chování numerického řešení počáteční úlohy pro testovací rovnici

$$y' = 0, y(0) = 0, x \in \langle 0, 1 \rangle. \qquad y' = \lambda y, y(0) = 1, x \in \langle 0, 1 \rangle.$$

$$y' = 1, y(0) = 0, x \in \langle 0, \infty \rangle. \qquad y' = \lambda, y(0) = 1, x \in \langle 0, \infty \rangle.$$

7. (2b.) Hermitův interpolační polynom má čtyři uzly. Počet podmínek v nich zadaných je postupně 2, 4, 4 a 1. Vyberte správné tvrzení:

Hodnoty první derivace jsou předepsány ve čtyřech uzlech.

Hodnoty druhé derivace jsou předepsány ve dvou uzlech.

Hodnota třetí derivace je předepsána v jednom uzlu.

Funkční hodnoty jsou předepsány ve čtyřech uzlech.

8. (1b.) Počítáme-li integrál na intervalu $\langle a, b \rangle$ různém od $\langle -1, 1 \rangle$, pak použití některé Gaussovy kvadraturní formule

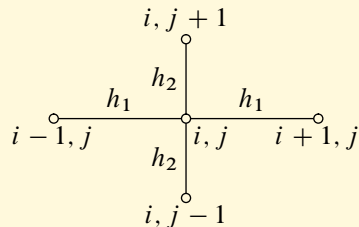
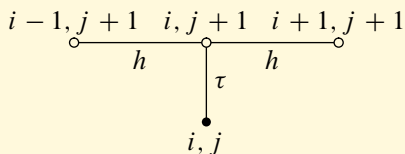
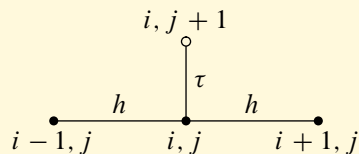
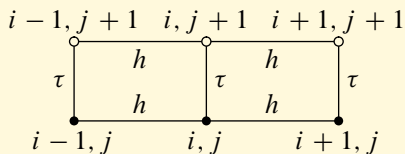
není vhodné, protože se dopouštíme velkých chyb.

není možné.

vyžaduje nejprve provedení lineární substituce, která interval $\langle a, b \rangle$ převede na interval $\langle -1, 1 \rangle$.

vyžaduje rozdělení intervalu $\langle a, b \rangle$ na vhodné podintervaly.

9. (1b.) Která šablona odpovídá ryze implicitnímu diferenčnímu schématu pro smíšenou úlohu pro rovnici vedení tepla?



10. (1b.) Vyberte tvrzení, které platí o explicitní Eulerově metodě.

Tato metoda dává poměrně dobré výsledky.

Tato metoda patří mezi více krokové metody.

Tato metoda je ideální pro nalezení hodnot y_1, y_2, \dots, y_{k-1} do k -krokových metod.

Tato metoda je velmi jednoduchá, ale nepřesná.

11. (2b.) Vyberte nepravdivé tvrzení o úloze numerické derivace:

Je to dobře podmíněná úloha.

Je to špatně podmíněná úloha.

Používáme při ní náhradu
interpoláčnım polynomem.

Derivaci numericky počítat nelze.

Správně zodpovězené otázky:

Získané body:

Procento úspěšnosti:

Literatura

- [1] Anděl, J. *Statistické metody*. 4. vydání. Praha: MATFYZPRESS, 2007. 300 s. ISBN 80-7378-003-8.
- [2] Babuška, I., Práger, M., Vitásek, E. *Numerické řešení diferenciálních rovnic*. 1. vydání. Praha: SNTL, 1964. 240 s.
- [3] Bachvalov, N. S. *Číslennyje metody*. 2. vydání. Moskva: Nauka, 1975. 632 s.
- [4] Berezin, I. S., Židkov, N. P. *Metody vyčíslenij I, II*. 2. vydání. Moskva: Gosudarstvennoje izdatel'stvo fiziko-matematičeskoj literatury, 1962. 464+640 s.
- [5] Blaheta, R. *Matematické modelování a metoda konečných prvků* [online]. Skriptum. 1. vydání. Ostrava: VŠB, 2012, vi+123 s. Dostupné z <http://mi21.vsb.cz/modul/matematicke-modelovani-metoda-konecnnych-prvku-numericke-metody-2> [cit. 2019-06-06].

- [6] Brandner, M., Egermaier, J., Kopincová, H. *Numerické metody pro řešení evolučních parciálních diferenciálních rovnic* [online]. Skriptum. 1. vydání. Ostrava: VŠB, 2012. vi+125 s. Dostupné z <http://mi21.vsb.cz/modul/numericke-metody-pro-reseni-evolucnich-parcialnich-diferencialnich-rovnic> [cit. 2019-06-06].
- [7] Butcher, J. C. *Numerical Methods for Ordinary Differential Equations*. 2nd edition. Chichester: John Wiley & Sons Ltd, 2008. xx+463 pp. ISBN 978-0-470-72335-7.
- [8] Collatz, L. *Funkcionální analýza a numerická matematika*. 1. vydání. Praha: SNTL, 1970. 420 s.
- [9] Čermák, L. *Numerické metody II. Diferenciální rovnice*. Skriptum. 2. upravené vydání. Brno: Vysoké učení technické v Brně, FSI, 2010. 136 s. ISBN 978-80-214-4110-1.
- [10] Čermák, L. *Vybrané statě z numerických metod* [online]. Skriptum. Brno: Vysoké učení technické v Brně, FSI, 2013. 97 s. Dostupné z <http://mathonline.fme.vutbr.cz/Numericke-metody-I/sc-1150-sr-1-a-141/default.aspx> [cit. 2019-06-06].

- [11] Čermák, L. *Numerické metody pro řešení diferenciálních rovnic* [online]. Skriptum. Brno: Vysoké učení technické v Brně, FSI, 2015. 81 s. Dostupné z <http://mathonline.fme.vutbr.cz/Numericke-metody-II/sc-1246-sr-1-a-263/default.aspx> [cit. 2019-06-06].
- [12] Čermák, L. *Algoritmy metody konečných prvků* [online]. Skriptum. Brno: Vysoké učení technické v Brně, FSI, 2005. 107 s. Dostupné z <http://mathonline.fme.vutbr.cz/Numericke-metody-III/sc-1151-sr-1-a-142/default.aspx> [cit. 2019-06-06].
- [13] Čermák, L., Hlavička, R. *Numerické metody*. Skriptum. 2. vydání. Brno: Vysoké učení technické v Brně, FSI, 2008. 110 s. ISBN 978-80-214-3752-4.
- [14] Děmidovič, B. P., Maron, I. A. *Základy numerické matematiky*. 1. vydání. Praha: SNTL, 1966. 724 s.
- [15] Došlá, Z., Došlý, O. Skriptum. *Metrické prostory*. 4. vydání. Brno: Přírodovědecká fakulta Masarykovy univerzity, 2016. viii+91 s. ISBN 978-80-210-8357-8.
- [16] Došlá, Z., Kuben, J. *Diferenciální počet funkcí jedné proměnné*. Skriptum. 2. vydání. Brno: Masarykova univerzita v Brně, 2012. vi+209 s. ISBN 978-80-210-5814-9.

- [17] Faddějev, D. K., Faddějevová, V. N. *Numerické metody lineární algebry*. 2. vydání (1. české). Praha: SNTL, 1964. 684 s.
- [18] Fiedler, M. *Speciální matice a jejich použití v numerické matematice*. 1. vydání. Praha: SNTL, 1981. 272 s.
- [19] Hairer, E., Nørsett, S. P., Wanner, G. *Solving Ordinary Differential Equations I. Nonstiff Problems*. 2nd edition, corrected 3rd printing 2008. Berlin: Springer-Verlag, 1993. xvi+528 pp. ISBN 978-3-540-56670-0.
- [20] Hairer, E., Wanner, G. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. 2nd edition, corrected printing 2002. Berlin: Springer-Verlag, 1996. xvi+614 pp. ISBN 978-3-540-60452-5.
- [21] Hairer, E., Lubich, Ch., Wanner, G. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. 2nd edition. Berlin: Springer-Verlag, 2006. xviii+644 pp. ISBN 978-3-540-30663-4.
- [22] Horová, I., Zelinka, J. *Numerické metody*. Skriptum. 2. rozšířené vydání. Brno: Přírodovědecká fakulta Masarykovy univerzity, 2008. vi+293 s. ISBN 978-80-210-3317-7. Dostupné z <http://www.math.muni.cz/~zelinka/dokumenty/numerika.pdf> [cit. 2019-06-06].

- [23] Hosea M. E. A new recurrence for computing Runge-Kutta truncation error coefficients. *SIAM J. Numer. Anal.*, Vol. 32, No 6 (1995), pp. 1989–2001.
- [24] Hošková, Š., Kuben, J. *Integrální počet funkcí jedné proměnné*. Skriptum. 1. vydání. Brno: Vojenská akademie v Brně, 2004. vi+197 s. ISBN 80-85960-75-3.
- [25] Hošková, Š., Kuben, J., Račková, P. *Integrální počet funkcí jedné proměnné* [online]. Skriptum. 1. vydání. Ostrava: VŠB–TU, 2006. vi+220 s. ISBN 80-248-1191-X. Obrazková verze 361 s. ISBN 978-80-248-1305-9. Dostupné z <http://homel.vsb.cz/~sla64/cd/> [cit. 2019-06-06].
- [26] Householder, A. S. *The Theory of Matrices in Numerical Analysis*. 1st edition. New York: Blaisdell Publishing Company, 1964. xiv+257 pp.
- [27] Isaacson, E., Keller, H. B. *Analysis of Numerical Methods*. Dover edition. New York: Dover Publications, Inc., 1994. xvi+541 pp. ISBN 0-486-68029-0.
- [28] Jevický, J., Kovařík, P. *Numerické metody algebry*. Skriptum. 1. vydání. Brno: VAAZ, 1986. 124 s. Dostupné v knihovně UO pod číslem S 2396.
- [29] Jevický, J., Kovařík, P. *Numerické metody analýzy*. Skriptum. 1. vydání. Brno: VAAZ, 1987. 170 s. Dostupné v knihovně UO pod číslem S 2397.

- [30] Kantorovič, L.V., Akilov, G.P. *Funkcional'nyj analiz*. 3. vydání. Moskva: Nauka, 1984. 703 s.
- [31] Kornejčuk, N.P. *Splajny v teorii približenija*. 1. vydání. Moskva: Nauka, 1984. 352 s.
- [32] Kosmák, L., Potůček, R. *Metrické prostory*. 1. vydání. Praha: Academia, 2004. 98 s. ISBN 80-200-1202-8.
- [33] Kuben, J. *Diferenciální počet funkcí jedné proměnné*. Skriptum. 1. vydání. Brno: Univerzita obrany, 2015. viii+333 s. ISBN 978-80-7231-991-6. Dostupné v knihovně UO pod číslem S 3854.
- [34] Kuben, J. *Obyčejné diferenciální rovnice*. Skriptum. 4. vydání. Brno: VA, 2000. vi+124 s. Dostupné v knihovně UO pod číslem S 18C.
- [35] Kuben, J., Račková, P. *Numerické metody*. Skriptum. 1. vydání. Brno: Univerzita obrany, 2016. viii+245 s. ISBN 978-80-7231-373-0. Dostupné v knihovně UO pod číslem S 3876.
- [36] Kufner, A. *Geometrie Hilbertova prostoru*. Praha: SNTL, 1973. 248 s.

- [37] Lambert J.D. *Numerical Methods for Ordinary Differential Systems: the Initial Value Problem*. Chichester: John Wiley & Sons Ltd, 1993. x+294 pp. ISBN 978-0-471-92990-1.
- [38] Likeš, J., Machek, J. *Matematická statistika*. 1. vydání. Praha: SNTL, 1983. 180 s. Dostupné v knihovně UO pod číslem S 2670/11.
- [39] Marčuk, G.I. *Metody numerické matematiky*. 1. vydání. Praha: Academia, 1987. 528 s.
- [40] Míka, S. *Numerické metody algebry*. 1. vydání. Praha: SNTL, 1982. 176 s. Dostupné v knihovně UO pod číslem S 2670/4.
- [41] Pospíšil, L., Vondrák, V. *Numerické metody I* [online]. Skriptum. 1. vydání. Ostrava: VŠB, 2011. vii+184 s. Dostupné z <http://mi21.vsb.cz/modul/numericke-metody-1> [cit. 2019-06-06].
- [42] Příkryl, P. *Numerické metody analýzy*. 1. vydání. Praha: SNTL, 1985. 192 s. Dostupné v knihovně UO pod číslem S 2670/24.
- [43] Ralston, A. *Základy numerické matematiky*. 1. vydání. Praha: Academia, 1973. 636 s.

- [44] Rektorys, K. *Metoda časové diskretizace a parciální diferenciální rovnice*. 1. vydání. Praha: SNTL, 1985. 364 s.
- [45] Rektorys, K. *Variační metody v inženýrských problémech a v problémech matematické fyziky*. 6. vydání (2. české opravené). Praha: Academia, 1999. 604 s. ISBN 80-200-0714-8.
- [46] Samarskij, A. A. *Vvedenije v čislennyje metody*. 1. vydání. Moskva: Nauka, 1982. 272 s.
- [47] Samarskij, A. A. *Teorija raznostnych schem*. 2. vydání, upravené. Moskva: Nauka, 1983. 616 s.
- [48] Samarskij, A. A., Gulin, A. V. *Ustojčivosť raznostnych schem*. 1. vydání. Moskva: Nauka, 1973. 416 s.
- [49] Samarskij, A. A., Nikolajev, J. S. *Numerické řešení velkých řídkých soustav*. 1. vydání. Praha: Academia, 1984. 600 s.
- [50] Shampine, L. F. *Numerical solution of ordinary differential equations*. New York: Chapman & Hall, 1994. x+484 pp. ISBN 978-0-412-05151-7.

- [51] Steffensen, J. F. *Interpolation*. Baltimore: The Williams & Wilkins Company, 1927. x+248 pp.
- [52] Stoer, J., Bulirsch, R. *Introduction to Numerical Analysis*. 3rd edition. New York: Springer, 2010. xvi+752 pp. ISBN 978-1-4419-3006-4.
- [53] Szegő, G. *Orthogonal polynomials*. 4rd edition. New York: AMS, 2003. 432 pp. ISBN 978-0-8218-1023-5. Překlad 2. vydání do ruštiny Moskva: Gosudarstvennoje izdatel'stvo fiziko-matematičeskoj literatury, 1962. 500 s.
- [54] Taylor, A. E. *Úvod do funkcionální analýzy*. 1. vydání. Praha: Academia 1973. 412 s.
- [55] Varga, R. S. *Matrix Iterative Analysis*. 2nd revised and expanded edition. Berlin: Springer 2000. x+358 pp. ISBN 3-540-66321-5.
- [56] Vitásek, E. *Numerické metody*. 1. vydání. Praha: SNTL, 1987. 561 s.
- [57] Vitásek, E. *Základy teorie numerických metod pro řešení diferenciálních rovnic*. 1. vydání. Praha: Academia, 1994. 412 s. ISBN 80-200-0281-2.
- [58] Volkov, Je. A. *Čislennyje metody*. 1. vydání. Moskva: Nauka, 1982. 256 s.
- [59] Zavjalov, Ju. S., Kvasov, B. I., Mirošničenko, V. L. *Metody splajn-funkcij*. 1. vydání. Moskva: Nauka, 1980. 352 s.

Další zdroje dostupné z Internetu

- [60] *Arbitrary-precision arithmetic* [online]. Dostupné z http://en.wikipedia.org/wiki/Arbitrary-precision_arithmetic [cit. 2019-06-06].
- [61] *Landauovy symboly* [online]. Dostupné z http://en.wikipedia.org/wiki/Big_O_notation [cit. 2019-06-06].
- [62] Čermák, L., Hlavička, R. *Numerické metody* [online]. Obrazovková verze kapitoly 1 z [13], 2006. Dostupné z <http://mathonline.fme.vutbr.cz/UploadedFiles/231.pdf> [cit. 2019-06-06].
- [63] Čermák, L., Hlavička, R. *Numerické metody* [online]. Obrazovková verze kapitoly 2 z [13], 2006. Dostupné z <http://mathonline.fme.vutbr.cz/UploadedFiles/233.pdf> [cit. 2019-06-06].
- [64] Hadjidimos, A. *Successive overrelaxation (SOR) and related methods*. Journal of Computational and Applied Mathematics 123 (2000), str. 177–199. Dostupné z <http://www.sciencedirect.com/science/article/pii/S0377042700004039> [cit. 2019-06-06].

- [65] *Differential Equations* [online]. Dostupné z <http://www.mymathlib.com/diffeq/> [cit. 2019-06-06].
- [66] Epperson, J. F. *On the Runge Example* [online]. Dostupné z http://www.maa.org/sites/default/files/pdf/upload_library/22/Ford/Epperson329-341.pdf [cit. 2019-06-06].
- [67] *Embedded Runge-Kutta Methods* [online]. Dostupné z http://www.mymathlib.com/diffeq/embedded_runge_kutta/ [cit. 2019-06-06].
- [68] Feistauer, M. *Základy numerické matematiky* [online]. Skriptum. Praha, MFF UK. 56 s. Dostupné z <http://www.karlin.mff.cuni.cz/~feist/ZNM-scripta.pdf> [cit. 2019-06-06].
- [69] Felcman, J. *Numerická matematika* [online]. Skriptum. Praha, MFF UK, KNM Press, 2017. 52 s. Dostupné z <http://www.karlin.mff.cuni.cz/~felcman/nm.pdf> [cit. 2019-06-06].
- [70] *Churchova-Turingova teze* [online]. Dostupné z http://en.wikipedia.org/wiki/Church-Turing_thesis [cit. 2019-06-06].

- [71] *IEEE Standard for Floating-Point Arithmetic (IEEE 754)* [online]. Dostupné z http://en.wikipedia.org/wiki/IEEE_754 [cit. 2019-06-06].
- [72] *List of Runge-Kutta methods* [online]. Dostupné z http://en.wikipedia.org/wiki/List_of_Runge-Kutta_methods [cit. 2019-06-06].
- [73] *LSodar* [online]. Dostupné z http://help.scilab.org/docs/5.4.1/en_US/LSodar.html [cit. 2019-06-06].
- [74] *Maple worksheets on the derivation of order 8 Runge-Kutta schemes* [online]. Dostupné z http://www.peterstone.name/Maplepgs/RKcoeff_8.html [cit. 2019-06-06].
- [75] *Not a number* [online]. Dostupné z <http://en.wikipedia.org/wiki/NaN> [cit. 2019-06-06].
- [76] Verner, J. *A Retrospective Survey on Deriving Explicit Runge-Kutta Pairs* [online]. Dostupné z <http://www.fields.utoronto.ca/programs/scientific/11-12/SciCADE2011/presentations/Verner.pdf> [cit. 2019-06-06].

- [77] *Domovská stránka Jima Vernerera* [online]. <http://people.math.sfu.ca/~jverner/> [cit. 2019-06-06].
- [78] Yu, Y. *Rosenbrock Methods* [online]. Dostupné z http://www.cfm.brown.edu/people/jansh/page5/page10/page40/assets/Yu_Talk.pdf [cit. 2019-06-06].

Rejstřík

A

- algorithmus, 34
 - numerický, 19, 35
 - numericky stabilní, 35, 250
- aproximace
 - čísla, 20
 - interpolační, 306
 - kořenu, 116
 - metodou nejmenších čtverců, 306

B

- bariéra Dahlquistova
 - druhá, 632
 - první, 630

C

- cifra platná, 22

Č

- čárka pohyblivá
 - hardwarová, 28
 - softwarová, 28

číslo

matice vlastní, 60

podmíněnosti

matice, 64, 253

úlohy, 30

strojové, 27

D

diference

dopředná, 324

poměrná, 319

zobecněná, 356

vpřed, 324

vzad, 325

zpětná, 325

doplňěk ortogonální, 76

E

exponent, 27

extrapolace, 305

F

formule

Gaussovy-Legendrovy, 509

kvadrurní, 468

Gaussovy, 503

složené, 494

symetrická, 474

Newtonovy-Cotesovy, 476

otevřené, 484

složené, 494

uzavřené, 476

Rombergova, 520

H

hodnota absolutní, 39

Ch

chod GEM

 přímý, 209

 zpětný, 214

chyba

 aproximace

 absolutní, 20

 relativní, 21

 kvadrurní formule, 468

 matematického modelu, 20

 metody pro řešení ODR

 globální diskretizační, 597, 626

 lokální, 597

 lokální diskretizační, 596, 626

 numerické úlohy, 20

 ve vstupních datech, 20

 zaokrouhlovací, 20

I

interpolace, 305

 iterovaná, 317

interval absolutní stability, 602

K

konvergence

 kvadratická, 122

 lineární, 122

 metody pro řešení ODR, 600, 630

 superlineární, 122

 v normě, 47, 52

korektor, 636

kořen

rovnice, 114

soustavy rovnic, 170

kritérium Sylvestrovo, 226

M

mantis normalizovaná, 27

matice

dobře podmíněná, 65

Gramova, 83, 394

hermitovská, 63

chybová, 250

ireducibilní, 273

iterační, 259

konjugovaná, 63

pásová, 228

ryze diagonálně dominantní

řádkově, 223

sloupcově, 223

řídká, 227

symetrická, 60, 225

indefinitní, 227

negativně definitní, 227

negativně semidefinitní, 227

pozitivně definitní, 60, 225

pozitivně semidefinitní, 60, 226

špatně podmíněná, 65

metoda

A-stabilní, 602, 632

bisekce, 125

D-stabilní, 629

Eulerova

explicitní, 590

implicitní, 593

- Gaussova eliminační, 207
- s částečným výběrem hlavních prvků, 215
 - s úplným výběrem hlavních prvků, 218
- Gaussova-Seidelova, 270
- Gillova, 620
- Heunova, 615
- Choleského, 246
- Jacobiova, 266
- konzistentní, 627
- Kuttova, 616
- L-stabilní, 613
- LU rozkladu, 233
- s částečným výběrem hlavních prvků, 240
- nejmenších čtverců, 390
- Newtonova, 154, 175
- prosté iterace, 143, 173
- půlení intervalu, 125
- regula falsi, 131
- relaxační, 275
- Rungeho-Kutty klasická, 619
- sečen, 161
- sítí, 691
- SOR, 276
- stabilní ve smyslu Dahlquistova, 629
- startovací, 123
- tečen, 154
- tětiv, 131
- zpřesňující, 123
- metody
- Adamsovy-Bashforthovy, 634
 - Adamsovy-Moultonovy, 635

prediktor-korektor, 636
pro řešení ODR jedнокrokové, 586
 explicitní, 588
 implicitní, 592
pro řešení ODR vícečrokové, 586
pro řešení ODR vícečrokové lineární,
 623
 explicitní, 624
 implicitní, 625
pro řešení systémů lineárních rovnic
 finitní, 206
 iterační, 206, 257
 přímé, 206, 207
Rungeho-Kutty, 606
 vnořené, 621
zpětného derivování, 638
místo desetinné platné, 22

model matematický, 18

N

norma

 maticová, 50

 Frobeniova, 51

 Hilbertova-Schmidtova, 52

 indukovaná vektorovými norma-
 mi, 56

 maximální, 51

 operátorová, 56

 přidružená k vektorovým normám,
 56

 Schurova, 52

 součtová, 51

 souhlasná s vektorovou normou,
 54

- submultiplikativní, 54
- vektorová, 41
 - eukleidovská, 42
 - indukovaná skalárním součinem, 70
 - maximální, 42
 - součtová, 42
- normy
 - ekvivalentní, 47, 52
- O**
- oblast absolutní stability, 602, 632
- P**
- pivot, 209
- podmínka
 - okrajová, 701, 712, 729
 - počáteční, 712, 729
 - zastavovací, 119, 171, 261
- podmínky
 - Fourierovy, 139, 157
- poloměr matice spektrální, 67, 260
- polynom
 - interpolační, 307
 - Hermitův, 338
 - Lagrangeův, 309
 - Newtonův, 318
 - Newtonův vpřed, 329
 - Newtonův vzad, 330
 - Lagrangeův, 309
 - zobecněný, 344
 - Legendrův, 505
- pravidla zaokrouhlovací, 24
- pravidlo
 - Boolovo, 480

- lichoběžníkové, 480
 - složené, 498
- Milnovo, 488
- obdélníkové, 488
 - složené, 496
- otevřené lichoběžníkové, 488
- Simpsonovo, 480
 - složené, 500
- Simpsonovo 3/8, 480
- prediktor, 636
- problém reálný, 18
- průmět pravouhlý na podprostor, 77
- prvek hlavní, 209

- R**
- reziduum, 251

- rovnice
 - diferenciální
 - obyčejná, 570
 - parciální, 690
 - tuhá (stiff), 640
 - difuzní, 712
 - Laplaceova, 701
 - Poissonova, 701
 - síťová, 692, 694
 - vedení tepla, 712
 - vlnová, 729
- rozklad matice
 - Choleského, 244
 - LU, 229
 - s částečným výběrem hlavních prvků, 237

Ř

řád

- aproximace diferenčního schématu, 697
- iterační metody, 122
- konvergence diferenčního schématu, 697
- konvergence posloupnosti, 122
- metody pro řešení ODR, 599, 626

S

- separace kořenu, 116
- schéma diferenční, 694
 - aproximující PDR, 697
 - Crankovo-Nicolsonové, 720
 - konvergentní, 697
 - korektní, 696
 - pro Poissonovu rovnici, 703

- pro rovnici vedení tepla, 716
- pro vlnovou rovnici, 734
- stabilní, 696

sít, 691

- součin skalární, 69
- standardní, 72

soustava rovnic normální, 82, 394

splajn, 367

kubický interpolační, 368

- not a knot, 371
- periodický, 371
- přirozený, 371
- úplný, 371

stupeň

- metody Rungeho-Kutty, 606
- přesnosti kvadrurní formule, 469

symbol O , 36

system čísel s pohyblivou řádovou čárkou,
26

Š

šablona, 692
 pro Poissonovu rovnici, 702
 pro rovnici vedení tepla, 715
 pro vlnovou rovnici, 732

T

tabulka Butcherova, 609
 teze Churchova-Turingova, 34

U

úloha
 Dirichletova, 701
 dobře podmíněná, 30
 korektní, 29
 numerická, 18

špatně podmíněná, 30

uzel, 583, 691
 hraniční, 691
 vnitřní, 691

V

vektory
 kolmé, 74
 ortogonální, 74
 vzorec interpolační
 Besselův, 334
 Everettův, 334
 Gaussův, 334
 Steffensenův, 335
 Stirlingův, 334

Z

zápis semilogaritmický, 26

Název: Numerické metody
Autoři: doc. RNDr. Jaromír Kuben, CSc.
PhDr. Pavlína Račková, Ph.D.
Vydavatel: Univerzita obrany v Brně
Tisk: Univerzita obrany v Brně
Počet stran: 817
Rok vydání: 2019
Vydání: první

ISBN 978-80-7582-092-1

Publikace neprošla jazykovou úpravou.