Numerické metódy

Mgr. Jela Babušíková, PhD.
Dr.rer.nat.habil. RNDr. Marián Slodička, PhD.
RNDr. Juraj Weisz, CSc.

Obsah

Predslov

Človek v nekonečnom procese poznávania prechádza postupne od pozorovania nových javov k ich štúdiu. Snaží sa pochopiť ich zákonitosti, precízne ich sformulovať v matematickej reči a číselne ich kvantifikovať. Tento proces modelovania je komplexný a interdisciplinárny. Sklbuje v sebe prácu i dôvtip špecialistov z rôznych odvetví i vedných disciplín. Numerická matematika je pevnou súčasťou tohto procesu. Nestojí však ani na jeho začiatku ani na konci. Treba ju chápať ako jedno ohnivko reťaze vedúcej od problému k jeho riešeniu. Logicky by mala nasledovať až po kvalitatívnej i kvantitatívnej analýze daného problému, no v skutočnosti tomu tak nemusí byť. Často sa nové neznáme úlohy riešia najprv numericky a až potom sa vytvorí teoretický základ pre ich ucelené štúdium. Veľakrát sa numerická matematika chápe ako súhrn vzorcov i postupov na číselné riešenie daných úloh a to podotknime bez toho, aby sa pochopili vzájomné súvislosti numerických metód so samotnou podstatou problému. Takáto nevedomosť môže viesť k tomu, že mnohokrát overená výpočtová schéma zlyhá na novej úlohe a číselným výstupom budú fyzikálne nereálne hodnoty. Z tohto dôvodu sa na numerickú matematiku nemôžeme pozerať ako na *čiernu skrinku*, ktorá zahrkoce a vydá výsledky, ak jej dáme patričné vstupné dáta. Je preto potrebné zoznámiť sa už i s klasickými približnými metódami a správne sa medzi nimi orientovať. Treba ich pochopiť, veď v nich sa často skrýva ideové bohatstvo, ktoré sa po istom zovšeobecnení dá aplikovať na riešenie nových oveľa zložitejších problémov. Ďalej si treba uvedomiť, že o vhodnosti resp. úspešnosti danej numerickej metódy pri riešení konkrétneho problému rozhodujú i iné činitele ako napr. použitá dátová štruktúra, architektúra počítača atď.

Nechceli sme podať výklad numerických metód ako suchý receptár, ale v mnohých prípadoch sme sa snažili o ich vysvetlenie z hľadiska matematickej analýzy či lineárnej algebry. V texte sa okrem tvrdení a ich dôkazov nachádzajú i riešené ilustračné príklady. Úloh, ktorých riešenie je ponechané na čitateľa, pravdupovediac nie je veľa. Naše skúsenosti ukazujú vhodnosť kombinácie praktických a teoretických cvičení. Pri tabuli si študent môže overiť, či zvládol teoretický základ podaný na prednáške a pri praktickom cvičení zas môže do chuti experimentovať s rôznymi príkladmi. Zároveň sa pritom využíva hravosť a chuť mladej generácie pracovať pri počítači. Výsledky príkladov sa dajú vizuálne znázorniť, čo znásobuje efektívnosť práce i štúdia. Veď predsa dobrý obrázok je oveľa presvedčivejším a názornejším agrumentom ako tabuľka čísel. V tomto smere majú dnešní študenti oveľa lepšie možnosti ako sme ich svojho času mali my. Je to podmienené obrovským rozmachom výpočtovej techniky v posledných rokoch. Zámerne sme neuvádzali naprogramované algoritmy. Tu sme prenechali voľnosť cvičiacemu vo voľbe príkladov i samotného programovacieho jazyka. Vhodnými alternatívami sú MATHEMATICA alebo MATLAB. Sú rozšírené, existujú k nim príručky a skrývajú v sebe už implementované obrovské množstvo algoritmov.

Publikácia je napísaná na základe skúseností autorov z prednášok i cvičení z numerických metód. Svojim obsahom predstavuje základný kurz numerickej matematiky. Je venovaná študentom druhých resp. tretích ročníkov rôznych študijných odborov univerzitného smeru. Predpokladá sa znalosť základného kurzu matematickej analýzy a lineárnej algebry. Súčasne predpokladáme aspoň základné vedomosti z teórie algoritmov a programovania kvôli lepšiemu pochopeniu uvádzaných algoritmov. Látka je rozdelená do desiatich kapitol tak, že jednotlivé časti sa dajú študovať nezávisle na sebe. Na začiatku sa čitateľ oboznámi so základnými pojmami z reprezentácie čísel a počítačovej aritmetiky. Tretia kapitola sa zaoberá riešením rôznych typov rekurentných relácií. V ďalších dvoch častiach sa pojednáva o interpolácii funkcie polynómom resp. splajnom a o aproximácii pomocou metódy najmenších štvorcov alebo lineárneho krigingu. Šiesta kapitola opisuje základné iteračné metódy riešenia systému lineárnych algebraických rovníc. Potom prejdeme k riešeniu nelineárnych rovníc typu $f(\boldsymbol{x}) = 0$. Ôsma kapitola sa zaoberá numerickým výpočtom integrálu pomocou Newtonových-Cotesových vzorcov a ich zložených tvarov. Posledné dve kapitoly vyžadujú isté vedomosti z teórie obyčajných i parciálnych diferenciálnych rovníc a sú venované základným technikám ich numerického riešenia.

Bratislava, december 1998 Autori

Kapitola 1

Úvod

Numerická matematika je nástrojom na riešenie matematických problémov, ktoré vznikajú pri riešení úloh v rôznych vedných a technických disciplínach. Obvyklý postup pri riešení reálnych problémov je približne nasledujúci:

1. Formulácia problému v reči príslušného odboru: (fyzika, chémia, biológia, ...).
2. Formulácia matematického modelu. Matematický model môže mať napríklad tvar hľadania minima funkcie, hľadania koreňa rovnice alebo systému rovníc, diferenciálnej rovnice atď.
3. Formulácia zjednodušeného matematického modelu (ak je pôvodný matematický model príliš komplikovaný, nahradíme ho jednoduchším - napr. nelineárnu rovnicu nahradíme lineárnou).
4. Numerické riešenie (obvykle nevieme nájsť presné) problému sformulovaného v bode 2 resp. 3. Musíme preto použiť nejakú metódu, ktorá nám umožní nájsť približné riešenie problému a prípadne odhadnúť chybu tohto približného riešenia. Tento problém vzniká vlastne už pri tých najjednoduchších úlohách. Napr. ak máme vypočítať hodnotu $y := \sin(x)$ použijeme rozvoj do Taylorovho radu s tým, že uvažujeme iba konečný (dostatočne veľký) počet sčítancov.
5. Implementácia numerickej metódy (voľba vhodného algoritmu, dátovej štruktúry, počítača, programovacieho jazyka).
6. Reprezentácia výsledkov - vizualizácia vypočítaného riešenia, interpretácia v reči pôvodného problému resp. príslušného odboru.

Po vykonaní fázy 6 sa môžeme vrátiť späť do niektorej z fáz 1 až 5 a modifikovať formuláciu, resp. riešenie problému.

Ozrejmime si situáciu na niekoľkých jednoduchých príkladoch.

Príklad 1.1 Vypočítajme objem Zeme.

Aby sme zostrojili presný matematický model, museli by sme presne poznať matematické vyjadrenie plochy, ktorá tvorí povrch Zeme. Potom by sme mohli integrovaním získať hodnotu objemu Zeme. To však nie je možné. Preto zvolíme zjednodušený model - teleso Zeme budeme považovať za guľu s polomerom $r = 6378$ km. Potom pre objem V dostávame $V = 4.\frac{1}{3}\pi r^3$.

Keď toto číslo budeme počítať (napr. na kalkulačke alebo počítači) dopustíme sa niekoľkých zjednodušení resp. nepresností: nepoznáme presne hodnotu π, číslo $\frac{1}{3}$ nahradíme číslom $0,333\ldots3$, násobenie čísel vystupujúcich vo výraze $4.\frac{1}{3}\pi r^3$ sa nevykoná presne, ale dostaneme iba zaokrúhlenú hodnotu. Nakoniec teda vznikne otázka, či je vypočítaná hodnota dostatočne presná na účely, na ktoré ju potrebujeme. □

Príklad 1.2 Vzťah medzi napätím U a prúdom I v elektrickom obvode je

$$I = a\left(\mathrm{e}^{bU} - 1\right),$$
$$c = dI + U,$$

pričom a, b, c, d sú dané známe konštanty. Dosadením prvej rovnice do druhej dostávame pre U nelineárnu rovnicu

$$c = ad\left(\mathrm{e}^{bU} - 1\right) + U,$$

napr.

$$12 = 14,3\left(\mathrm{e}^{2U} - 1\right) + U.$$

Túto rovnicu nevieme presne vyriešiť. Musíme použiť nejakú numerickú metódu (napr. Newtonovu metódu alebo metódu bisekcie). Dostávame

$$U \approx 0,299\ .$$

S niektorými inými modelmi sa zoznámime v nasledujúcich kapitolách. □

Tieto prednášky sa venujú predovšetkým bodu 4, t.j. budeme hovoriť o základných numerických metódach na riešenie niektorých matematických problémov.

K tomu, aby bol nejaký matematický problém numericky riešiteľný, je rozumné predpokladať (resp. vyžadovať), že jeho riešenie existuje (prípadne je jediné) a že je stabilné vzhľadom k dátam úlohy. Predpokladáme, že pojmy existencie a jednoznačnosti riešenia sú čitateľovi dostatočne známe.

Pojem stability si ozrejmíme najprv na niekoľkých príkladoch.

Príklad 1.3 Systém lineárnych algebraických rovníc

$$2x_1 + 6x_2 = 8$$
$$2x_1 + 6,000001x_2 = 8,000001$$

má presné riešenie vektor $(1, 1)$.

Systém lineárnych algebraických rovníc

$$2x_1 + 6x_2 = 8$$
$$2x_1 + 5,999999x_2 = 8,000001$$

má riešenie $(7, -1)$.

To znamená, že malá zmena v koeficientoch vyvolala veľkú zmenu riešenia. Dôvod je v tom, že priamky, ktorých reprezentáciou je príslušný systém (ich priesečník je riešením) sú "takmer rovnobežné". □

Príklad 1.4 Systém $\boldsymbol{Ax} = \boldsymbol{b}$, kde

$$A = \begin{pmatrix} 0,1 & -1 & 0 & 0 & 0 \\ 0 & 0,1 & -1 & 0 & 0 \\ 0 & 0 & 0,1 & -1 & 0 \\ 0 & 0 & 0 & 0,1 & -1 \\ 0 & 0 & 0 & 0 & 0,1 \end{pmatrix}, \; b = \begin{pmatrix} 0,1 \\ -1 \\ 0,1 \\ -1 \\ 0,1 \end{pmatrix}$$

má riešenie $\boldsymbol{x} = (1; 0; 1; 0; 1)^T$.

Ak zvolíme $\tilde{\boldsymbol{b}} = (0,1; -1; 0,1; -1; 0,101)^T$ potom riešením systému $\boldsymbol{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$ je vektor $\tilde{\boldsymbol{x}} = (101; 10; 2; 0,1; 1,01)^T$. Malá zmena jedného prvku pravej strany vyvolala teda veľkú zmenu riešenia. □

Príklad 1.5 Uvažujme kvadratickú rovnicu

$$x^2 - 4x + 4 = 0.$$

Jej korene sú $r_1 = r_2 = 2$. Kvadratická rovnica $x^2 - 4x + 3,999 = 0$ má korene $r_1 = 2,01$ a $r_2 = 1,99$. Malá zmena koeficientov (o 10^{-3}) vyvolala teda veľkú zmenu riešenia (o 10^{-2}). □

Príklad 1.6 Nech postupnosť $\{z_n\}_{n=0}^{\infty}$ je daná rekurentným vzťahom

$$z_{n+1} = az_n,$$

pričom $a \geq 0$, $z_0 \geq 0$ sú dané. Ak

$$a > 1, \; z_0 = 0, \text{ potom } z_n = 0 \quad n = 1, 2, \ldots,$$

$$a > 1, \; z_0 \neq 0, \text{ potom } z_n = a^n z_0 \to \infty, \text{ a teda } \lim_{n\to\infty} z_n = \infty.$$

To znamená, že úloha vypočítať členy postupnosti z_n je nestabilná, lebo malá zmena vstupu spôsobí veľkú zmenu riešenia. Pritom v prípade $0 \leq a < 1$ tento jav nenastane. □

Príklad 1.7 Nech

$$f_1 : \mathbb{R} \to \mathbb{R}, \; f_1(x) = 0 \quad \forall x \in \mathbb{R},$$

$$f_2 : \mathbb{R} \to \mathbb{R}, \; f_2(x) = \varepsilon \sin(nx) \quad \forall x \in \mathbb{R},$$

kde ε, n sú nejaké kladné čísla. Platí $f_1' \equiv 0$, $f_2' = \varepsilon n \cos(nx)$.

Ak teraz ε je malé a n veľké (napr. $\varepsilon = 10^{-6}$, $n = 10^6$) potom rozdiel funcií f_1, f_2 je malý

$$\max |f_1 - f_2| \leq \varepsilon = 10^{-6},$$

ale rozdiel ich derivácií f_1', f_2' je veľký

$$\max |f_1' - f_2'| = |\varepsilon n \cos(nx)| = |\cos(nx)| = 1.$$

To znamená, že úloha "zderivovať danú funkciu" je v uvedenom zmysle vo všeobecnosti nestabilná. □

Príklad 1.8 Nech $f_i \in C([0,1])$, $i = 1, 2$ a nech

$$|f_1(x) - f_2(x)| \leq \varepsilon \qquad \forall x \in [0,1].$$

Potom

$$\left| \int_0^1 f_1(x)dx - \int_0^1 f_2(x)dx \right| \leq \int_0^1 |f_1(x) - f_2(x)|\, dx \leq \varepsilon.$$

To znamená, že malá zmena integrovanej funkcie vyvolá malú zmenu hodnoty určitého integrálu na intervale $[0,1]$. Úloha "integrovať spojitú funkciu definovanú na $[0,1]$" je teda stabilná. □

Pod pojmom *stabilita* rozumieme teda vlastnosť "malá zmena vstupných dát vyvolá malú zmenu riešenia". Problémy, ktoré nie sú stabilné si vyžadujú veľmi špeciálne numerické metódy (ak sa vôbec dajú uspokojivo riešiť).

Keďže v ďalšom texte sa budeme zaoberať predovšetkým numerickými metódami a ich realizáciou na počítači, bude vhodné najprv sa zaoberať otázkou reprezentácie reálnych čísel v počítači a počítačovou aritmetikou.

Kapitola 2

Reprezentácia čísel pri rôznych bázach

Obvyklý desiatkový zápis reálnych čísel používa číslice $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. Ľubovoľné celé (kladné) číslo môžeme pomocou týchto číslic zapísať v tvare

$$a_n a_{n-1} \dots a_1 a_0,$$

kde $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $i = 0, \dots, n$, $a_n \neq 0$. Napr.

$$3721 = 1 \cdot 10^0 + 2 \cdot 10^1 + 7 \cdot 10^2 + 3 \cdot 10^3.$$

Číslo z intervalu $(0, 1)$ zapisujeme v tvare

$$0, b_1 b_2 \cdots,$$

kde $b_i \in \{0, 1, \dots, 9\}$. Tento rozvoj však už môže byť nekonečný. Napr.

$$\begin{aligned} &0,254 = 2 \cdot 10^{-1} + 5 \cdot 10^{-2} + 4 \cdot 10^{-3} \\ &\tfrac{1}{3} = 0,333 \cdots = \sum_{k=1}^{\infty} 3 \cdot 10^{-k}. \end{aligned}$$

Každé reálne číslo môžeme teda zapísať v tvare

$$\pm \left(\sum_{k=0}^{n} a_k 10^k + \sum_{k=1}^{\infty} b_k 10^{-k} \right), \text{ kde } a_n \neq 0.$$

Skrátene $\pm a_n a_{n-1} \cdots a_1 a_0, b_1 b_2 b_3 \cdots$. Pritom prvá suma je celá časť a druhá suma je zlomková časť.

Doteraz sme reprezentovali každé číslo pri základe 10. To isté môžeme urobiť pri ľubovoľnom inom základe β (kde β je kladné celé číslo). Pri reprezentácii čísel v počítačoch sa používajú obvykle bázy $\beta = 2$, $\beta = 8$, $\beta = 16$.

Pri základe β reprezentujeme reálne číslo v tvare

$$\pm \left(\sum_{k=0}^{n} a_k \beta^k + \sum_{k=1}^{\infty} b_k \beta^{-k} \right),$$

kde β je kladné celé číslo a

$$\begin{aligned} &a_k \in \{0, 1, \dots, \beta - 1\} \quad k = 0, \dots n \\ &a_n \neq 0 \\ &b_k \in \{0, 1, \dots, \beta - 1\} \quad k = 1, \dots . \end{aligned}$$

Skrátene zapisujeme $[a_n \cdots a_0, b_1 b_2 \cdots]_\beta$.

Napr. pri báze $\beta = 2$

$$10,125 = [1010,001]_2 .$$

Jednoduchý algoritmus na prevod čísel medzi dvojkovou a desiatkovou sústavou si ukážeme na príklade.

Príklad 2.1 Vyjadrite číslo 254, 125 v dvojkovej sústave.

Riešenie: Najprv zoberieme celú časť a postupne delíme, dole píšeme zvyšok

$$\begin{array}{ccccccccc}
254:2 & = & 127 & :2= & 63 & :2= & 31 & :2 \\
 & & 0 & & 1 & & 1 & \\
 & = & 15 & :2= & 7 & :2= & 3 & :2 \\
 & & 1 & & 1 & & 1 & \\
 & = & 1 & :2= & 0 & & & \\
 & & 1 & & 1. & & &
\end{array}$$

Zvyšky zapíšeme "odzadu"

$$254 = [11111110]_2 .$$

Teraz vezmeme zlomkovú časť, vynásobíme 2 a celú časť "odtrhávame":

$$0,125 \times 2 = 0 \rfloor, 250 \times 2 = 0 \rfloor, 500 \times 2 = 1 \rfloor, 0.$$

Skrátene zapísané $0,125 = 0,001_2$.
Spolu teda $254,125 = [11111110,001]_2$. □

Príklad 2.2 Vyjadrite číslo $[1101,11]_2$ v desiatkovej sústave.

Riešenie:

$$[1101,11]_2 = 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = 13,75 \,. \qquad \square$$

Príklad 2.3 Vyjadrite číslo $\frac{1}{7}$ v desiatkovej sústave.

Riešenie:

$$\begin{array}{llllllll}
1:7 & = & 0,\overline{142857} & & & & & \\
1 \times 10 & \mapsto & 10:7=1 & & & & & \\
 & & 3 \times 10 \mapsto & 30:7=4 & & & & \\
 & & & 2 \times 10 \mapsto & 20:7=2 & & & \\
 & & & & 6 \times 10 \mapsto & 60:7=8 & & \\
 & & & & & 4 \times 10 \mapsto & 40:7=5 & \\
 & & & & & & 5 \times 10 \mapsto & 50:7=7 \\
 & & & & & & & 1
\end{array}$$

□

Príklad 2.4 Vyjadrite číslo $\frac{1}{7}$ v osmičkovej sústave.

Riešenie:

$$\begin{array}{lll}
1:7 & = & 0,\overline{1}_8 \\
1 \times 8 & \mapsto & 8:7=1 \\
 & & 1
\end{array}$$

□

Príklad 2.5 Vyjadrite číslo $\frac{1}{7}$ v šestnástkovej sústave.

Riešenie:

$$\begin{array}{lllll} 1:7 & = & 0,\overline{249}_{16} & & \\ 1\times 16 & \mapsto & 16:7=2 & & \\ & & 2\times 16 \mapsto & 32:7=4 & \\ & & & 4\times 16 \mapsto & 64:7=9 \\ & & & & 1 \end{array}$$

□

Príklad 2.6 Nech $k \in \mathbb{N}$, $a_i \in \{0,1,\ldots,9\}$, $1 \le i \le k$, $a_1 \ne 0$, $A = 0,\overline{a_1a_2\cdots a_k}$. Vyjadrite číslo A v tvare zlomku v desiatkovej sústave.

Riešenie: Platí

$$\begin{aligned} 0,\overline{a_1a_2\cdots a_k} &= A, \\ a_1a_2\cdots a_k,\overline{a_1a_2\cdots a_k} &= 10^k\,A. \end{aligned}$$

Odčítaním rovníc dostaneme

$$a_1a_2\cdots a_k = \left(10^k - 1\right)A,$$

teda

$$A = \frac{a_1a_2\cdots a_k}{10^k - 1}.$$

□

2.1 Počítačová reprezentácia reálnych čísel

Štandardný zápis reálneho čísla je celá časť, desatinná čiarka[1], zlomková časť, napr. $37,256$ alebo $0,000289$. Iný štandardný spôsob je tzv. normalizovaný vedecký zápis t.j. $0,37256 \times 10^2$ alebo $0,289 \times 10^{-3}$. (Prvá číslica za desatinnou čiarkou je nenulová.) Táto reprezentácia sa tiež nazýva normalizovaný zápis s pohyblivou rádovou čiarkou v decimálnom systéme

$$x = \pm 0,d_1\cdots d_t d_{t+1}\cdots \times 10^n,$$

kde $d_1 \ne 0$, $d_i \in \{0,1,2,3,4,5,6,7,8,9\}$, n je celé číslo. Ináč povedané

$$x = \pm r \times 10^n,\ \text{kde } \frac{1}{10} \le r < 1.$$

V systéme so základom β je to podobne

$$x = \pm r \times \beta^n, \quad \text{kde } \beta^{-1} \le r < 1,$$

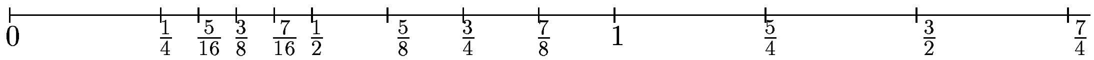
napr. pre $\beta = 2$

$$x = \pm r \times 2^n, \quad \text{kde } \frac{1}{2} \le r < 1.$$

Číslo r sa nazýva normalizovaná mantisa, n je exponent. V prípade $\beta = 2$ je $r_1 = 1$ a postupnosť $\{r_i\}$ je vlastne postupnosť bitov reprezentujúca dané číslo.

Keďže kapacita každého počítača je konečná, obrazom všetkých reálnych čísel je iba konečná množina - tzv. množina počítačových čísel. Táto množina vznikne tak, že

[1] V počítačoch sa používa desatinná bodka.



Obr. 2.1. Množina $\mathcal{M}(2,3,-1,1)$

uvažujeme čísla vo vedeckom zápise s mantisou dĺžky najviac t pri základe β, pričom exponent sa môže pohybovať v hraniciach $L \leq l \leq U$, kde $L, U \in \mathbb{Z}$.

Množina počítačových čísel je teda množina

$$\begin{aligned}\mathcal{M}(\beta,t,L,U) &= \left\{\pm(0,a_1a_2\cdots a_t)\times\beta^l; a_1\neq 0, a_i\in\{0,1,\ldots,\beta-1\}, L\leq l\leq U\right\}\\ &\cup\ \{0\}.\end{aligned}$$

(Napr.: $\beta = 2, t = 24, L = -126, U = 127$ alebo $\beta = 2, t = 63, L = -1022, U = 1023$.)

Príklad 2.7 Funkcia $n!$ rastie veľmi rýchlo. Už pre $n = 70$ nadobúda hodnotu väčšiu ako 10^{100} a to pre niektoré počítače znamená "overflow". Navrhnite postup ako vypočítať $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ pre hodnoty $n = 70,\ k = 35$.

Riešenie: Zrejme platí

$$\begin{aligned}\binom{70}{35} &= \frac{70\cdot 69\cdot 68\cdot\ldots\cdot 36}{1\cdot 2\cdot 3\cdot\ldots\cdot 35}\\ &= \frac{70}{35}\cdot\frac{69}{34}\cdot\frac{68}{33}\cdot\ldots\cdot\frac{36}{1}\\ &\approx 1,12186278\ E\ 20.\end{aligned}$$

Vypočítaná hodnota leží v pamäťovom rozsahu počítača. □

Príklad 2.8 Vypíšte explicitne a zakreslite na reálnej osi množinu $\mathcal{M}(2,3,-1,1)$.

Riešenie: (kladná časť- obrázok 2.1.) □

Podobná situácia je aj pre inú voľbu β, t, L, U. Vidíme teda, že množina $\mathcal{M}$ je "nepravidelne rozhádzaná" na číselnej osi.

Keďže množina reálnych čísel tvorí základ pri budovaní celej matematickej analýzy a algebry, snažíme sa každéme reálnemu číslu priradiť "čo najbližšie" číslo z množiny $\mathcal{M}$. Obvykle sa to robí jedným z nasledujúcich spôsobov:

$$x = (0,a_1\cdots a_ta_{t+1}\cdots)\times\beta^l \longmapsto (0,a_1\cdots a_t)\beta^l =: fl_t^{cut}(x)$$

alebo

$$x = (0,a_1\cdots a_ta_{t+1}\cdots)\times\beta^l \longmapsto fl_t^{round}(x),$$

pričom

$$fl_t^{round}(x) := \begin{cases} (0,a_1\cdots a_t)\beta^l & \text{ak } a_{t+1} < \beta/2 \\ (0,a_1\cdots a_t)\beta^l + \underbrace{(0,0\cdots 01)}_{t \text{ miest za des. čiarkou}}\beta^l & \text{ak } a_{t+1} \geq \beta/2. \end{cases}$$

Zobrazenie fl_t^{cut} sa nazýva *useknutie*, fl_t^{round} sa nazýva *zaokrúhlenie* na t desatinných miest. (Premyslite si, čo to znamená v prípade $\beta = 10, t = 3$.)

Pozrime sa teraz akej chyby sa dopustíme, ak nahradíme číslo x číslom $fl_t^{cut}(x)$, resp. $fl_t^{round}(x)$.

Veta 2.1 *Označme* $\varepsilon_1 = \frac{fl_t^{cut}(x)-x}{x}$, $\varepsilon_2 = \frac{fl_t^{round}(x)-x}{x}$. *Platí*

$$\begin{array}{rcl} -\beta^{-t+1} & \leq \; \varepsilon_1 & \leq 0, \\ -\frac{1}{2}\beta^{-t+1} & \leq \; \varepsilon_2 & \leq \frac{1}{2}\beta^{-t+1}. \end{array}$$

DÔKAZ: Urobíme pre ε_1, pričom uvažujeme $x > 0$. Označme $\gamma = \beta - 1$. Pre absolútnu chybu platí

$$\begin{aligned} x - fl_t^{cut}(x) &= (0,a_1\cdots a_t a_{t+1}\cdots)\beta^l - (0,a_1\cdots a_t)\beta^l \\ &= (0,0\cdots 0a_{t+1}a_{t+2}\cdots)\beta^l \leq (0,0\cdots 0\gamma\gamma\cdots)\beta^l \\ &= \gamma(0,0\cdots 0111\cdots)\beta^l = \gamma\beta^l(\beta^{-(t+1)} + \beta^{-(t+2)} + \beta^{-(t+3)} + \ldots) \\ &= \gamma\beta^l\beta^{-(t+1)}\frac{1}{1-\beta^{-1}} = \frac{(\beta-1)\beta^{l-t-1}}{\frac{\beta-1}{\beta}} \\ &= \beta^{l-t}. \end{aligned}$$

Pre relatívnu chybu teda dostávame

$$0 \leq \frac{x - fl_t^{cut}(x)}{x} \leq \frac{\beta^{l-t}}{(0,a_1a_2\cdots)\beta^l} \leq \frac{\beta^{l-t}}{(0,10\cdots)\beta^l} = \frac{\beta^{l-t}}{\beta^{-1}\beta^l} = \beta^{-t+1}.$$

Ostatné prípady sa dokážu analogicky. □

Poznámka 2.1 *Nie je pravda, že každé číslo, ktoré má konečnú mantisu pri nejakom základe má konečnú mantisu aj pri inom základe. Napr.*

$$0,1_{10} = 0,000110011001100\cdots{}_2\,,$$

$$\frac{1}{3} = 0,333333\cdots{}_{10} = 3^{-1} = 0,1_3.$$

To znamená, že ak zadáme na vstupe (napr. na terminál) číslo $0,1$, *toto číslo nebude v skutočnosti v počítači reprezentované presne!* □

2.2 Aritmetické operácie s počítačovými číslami

Na príklade si ukážeme sčítanie dvoch počítačových čísel pri základe $\beta = 10, t = 3$:

$$\begin{aligned} 253 + 0,1 &= 0,253 \times 10^3 + 0,1 \times 10^0 \\ &= 0,253 \times 10^3 + 0,0001 \times 10^3 \\ &= (0,253 + 0,0001) \times 10^3. \end{aligned}$$

Výsledok operácie bude teda

$$fl_3(253,1) = 253,$$

pričom fl_3 znamená buď fl_3^{cut} alebo fl_3^{round}.

Neplatí teda

$$x > 0, y > 0 \Rightarrow x + y > x.$$

Dôsledok 2.1 *Pre počítačové operácie neplatí asociatívny zákon. Napr. v počítačovej aritmetike pre* $t = 3$

$$\begin{aligned} (253 - 253) + 0,1 &= 0,1 \\ 253 + (-253 + 0,1) &= 0. \end{aligned}$$ □

Vo všeobecnosti ak ω značí jednu z aritmetických operácií $\{+, -, \times, :\}$ a $\hat{\omega}$ jej počítačovú verziu, tak

$$a \ \hat{\omega} \ b = fl_t(a \ \omega \ b)$$

(pričom a, b sú počítačové čísla).

Ak $a, b \in \mathbb{R}$, potom

$$a \ \hat{\omega} \ b = fl_t\left(fl_t(a) \ \omega \ fl_t(b)\right).$$

(Na vykonanie presnej operácie ω použijeme tzv. rezervný register.)

Zaujíma nás teraz chyba, ktorej sa dopustíme, ak urobíme operáciu $\hat{\omega}$ s aproximáciou operandov namiesto operácie ω s pôvodnými operandami.

Nech x_T, y_T sú presné čísla a nech x_A, y_A sú ich aproximácie. Pre nejaké "malé" ε, η platí teda

$$x_T = x_A + \varepsilon, \quad y_T = y_A + \eta.$$

Zrejme

$$\begin{aligned} x_T \ \omega \ y_T - x_A \ \hat{\omega} \ y_A &= x_T \ \omega \ y_T - x_A \ \omega \ y_A + x_A \ \omega \ y_A - x_A \ \hat{\omega} \ y_A \\ &= (x_T \ \omega \ y_T - x_A \ \omega \ y_A) + (x_A \ \omega \ y_A - fl_t(x_A \ \omega \ y_A)). \end{aligned}$$

Druhá zátvorka je vlastne iba chyba zaokrúhlenia, ktorú sme už analyzovali vo vete 2.1. Prvá zátvorka je chyba, ktorej sa dopustíme, ak nahradíme operandy x_T, y_T operandami x_A, y_A a aritmetickú operáciu počítame presne.

Budeme teraz skúmať (relatívnu) chybu pre jednotlivé aritmetické operácie.

NÁSOBENIE. Pre absolútnu chybu platí

$$\begin{aligned} x_T y_T - x_A y_A &= x_T y_T - (x_T - \varepsilon)(y_T - \eta) \\ &= x_T y_T - x_T y_T + \varepsilon y_T + \eta x_T - \varepsilon\eta \\ &= \varepsilon y_T + \eta x_T - \varepsilon\eta. \end{aligned}$$

Za predpokladu, že $\left|\frac{\varepsilon}{x_T}\right|, \left|\frac{\eta}{y_T}\right|$ sú "malé" pre relatívnu chybu dostávame

$$\begin{aligned} Rel\,(x_A y_A) &= \left|\frac{x_T y_T - x_A y_A}{x_T y_T}\right| = \left|\frac{\varepsilon y_T + \eta x_T - \varepsilon\eta}{x_T y_T}\right| \\ &\leq \left|\frac{\varepsilon}{x_T}\right| + \left|\frac{\eta}{y_T}\right| + \left|\frac{\varepsilon\eta}{x_T y_T}\right| \approx Rel\,(x_A) + Rel\,(y_A)\,. \end{aligned}$$

PREVRÁTENÁ HODNOTA. Pre absolútnu chybu platí

$$\frac{1}{x_T} - \frac{1}{x_A} = \frac{1}{x_T} - \frac{1}{x_T - \varepsilon} = \frac{x_T - \varepsilon - x_T}{x_T(x_T - \varepsilon)} = -\frac{\varepsilon}{x_T(x_T - \varepsilon)}$$

a pre relatívnu chybu (za predpokladu, že $Rel\,(x_A)$ je "malá")

$$\begin{aligned} Rel\left(\frac{1}{x_A}\right) &= \left|\frac{\frac{-\varepsilon}{x_T(x_T-\varepsilon)}}{\frac{1}{x_T}}\right| = \left|\frac{\varepsilon}{x_T - \varepsilon}\right| = \left|\frac{\varepsilon}{x_T}\right|\left|\frac{x_T}{x_T - \varepsilon}\right| \\ &= Rel\,(x_A)\,\frac{1}{1 - Rel\,(x_A)} \approx Rel\,(x_A)\,. \end{aligned}$$

SÚČET čísel s rovnakým znamienkom. Nech $x_T > 0,\ y_T > 0$ (rovnako by sme skúmali prípad $x_T < 0,\ y_T < 0$). Pre absolútnu chybu platí

$$(x_T + y_T) - (x_A + y_A) = (x_T + y_T) - (x_T - \varepsilon + y_T - \eta) = \varepsilon + \eta.$$

Pre relatívnu chybu

$$\begin{aligned} Rel\,(x_A + y_A) &= \left|\frac{\varepsilon + \eta}{x_T + y_T}\right| \leq \frac{|\varepsilon|}{x_T + y_T} + \frac{|\eta|}{x_T + y_T} \\ &\leq \frac{|\varepsilon|}{x_T} + \frac{|\eta|}{y_T} = Rel\,(x_A) + Rel\,(y_A)\,. \end{aligned}$$

Ak by sme nepredpokladali, že x_T, y_T majú rovnaké znamienka, potom by sme nemohli predchádzajúcu nerovnosť napísať, lebo ak $x_T > 0, y_T < 0$ potom

$$Rel\,(x_A + y_A) = \left|\frac{\varepsilon + \eta}{x_T + y_T}\right|,$$

pričom toto číslo môže byť ľuboľne veľké aj keď $|\varepsilon + \eta|$ je malé (v prípade $x_T \approx -y_T$).

Príklad 2.9 Nech

$$\begin{aligned} x_T &= 0,996, \qquad x_A = 1,0, \\ y_T &= -0,994, \quad y_A = -0,99\,. \end{aligned}$$

Potom pre relatívne chyby dostaneme

$$\begin{aligned} Rel\,(x_A + y_A) &= \left|\tfrac{-0{,}004-0{,}004}{0{,}996-0{,}994}\right| &&= \tfrac{0{,}008}{0{,}002} &&= 4, \\ Rel\,(x_A) &= \left|\tfrac{0{,}004}{0{,}996}\right| &&= \tfrac{1}{249} &&< 0,005, \\ Rel\,(y_A) &= \left|\tfrac{0{,}004}{0{,}994}\right| &&< 0,005\,. \end{aligned}$$

Teda relatívna chyba vstupných hodnôt je menej ako 0,5 % a relatívna chyba výstupných hodnôt je 400 % . □

2.3 Šírenie chyby pri sčitovaní

Nech $n \in \mathbb{N}$, $n \geq 2$ a $\{x_i\}_{i=1}^{n}$ sú dané reálne čísla v počítačovej reprezentácii pre dané β, t, t.j. $x_i = fl_t(x_i)$ pre všetky i, pričom kvôli skráteniu zápisu píšeme fl_t miesto fl_t^{round}. Našou úlohou je skúmať šírenie zaokrúhľovacej chyby pri sčitovaní.

Označme

$$\varphi_m(x_1, \ldots, x_n) := \sum_{\nu=1}^{m} x_\nu.$$

Sčitovanie je dané rekurentným predpisom [2]

$$\begin{aligned}
\varphi_1(x_1, \ldots, x_n) &= x_1, \\
\varphi_k(x_1, \ldots, x_n) &= \varphi_{k-1}(x_1, \ldots, x_n) + x_k.
\end{aligned}$$

Z vety 2.1 vyplýva nasledujúce vyjadrenie pre ľubovoľné číslo x

$$fl_t(x) = x(1+\varepsilon), \qquad \text{pre nejaké } |\varepsilon| \leq \tau := \frac{1}{2}\beta^{1-t}.$$

Takže môžme písať

$$\begin{aligned}
fl_t(\varphi_n(x_1, \ldots, x_n)) &= fl_t(fl_t(\varphi_{n-1}(x_1, \ldots, x_n)) + x_n) \\
&= (fl_t(\varphi_{n-1}(x_1, \ldots, x_n)) + x_n)(1+\varepsilon_n) \\
&= (fl_t(fl_t(\varphi_{n-2}(x_1, \ldots, x_n)) + x_{n-1}) + x_n)(1+\varepsilon_n) \\
&= ((fl_t(\varphi_{n-2}(x_1, \ldots, x_n)) + x_{n-1})(1+\varepsilon_{n-1}) + x_n)(1+\varepsilon_n) \\
&= (fl_t(\varphi_{n-2}(x_1, \ldots, x_n)) + x_{n-1})(1+\varepsilon_{n-1})(1+\varepsilon_n) + x_n(1+\varepsilon_n) \\
&= \ldots \\
&= x_1 \prod_{\mu=2}^{n}(1+\varepsilon_\mu) + \sum_{\nu=2}^{n} x_\nu \prod_{\mu=\nu}^{n}(1+\varepsilon_\mu) \\
&= x_1(1+\eta_2) + \sum_{\nu=2}^{n} x_\nu(1+\eta_\nu),
\end{aligned}$$

pričom sme označili *rezonančné faktory*

$$1 + \eta_\nu = \prod_{\mu=\nu}^{n}(1+\varepsilon_\mu).$$

Skúsme teraz odhadnúť veľkosť rezonančných faktorov. Platí

$$\begin{array}{ccccc}
-\tau & \leq & \varepsilon_\mu & \leq & \tau, \\
1-\tau & \leq & 1+\varepsilon_\mu & \leq & 1+\tau, \\
(1-\tau)^{n+1-\nu} & \leq & \prod_{\mu=\nu}^{n}(1+\varepsilon_\mu) & \leq & (1+\tau)^{n+1-\nu}, \\
(1-\tau)^{n+1-\nu} & \leq & 1+\eta_\nu & \leq & (1+\tau)^{n+1-\nu}.
\end{array}$$

[2] Uvedomte si, že sčitovanie je binárna operácia. Ak chceme sčítať viacero čísel, tak to musíme robiť postupne. Najprv sčítame prvé dve, potom k výsledku pripočítame ďalšie.

Z tohto odhadu vidíme, že rezonančné faktory $1+\eta_\nu$ môžu byť tým väčšie, čím je index ν menší. V praxi to znamená, že prvý sčítanec môže zanášať najväčšiu chybu. Z tohto hľadiska by bolo asi najlepšie, ak $x_1, \ldots, x_n$ najprv usporiadame podľa veľkosti tak, že začneme číslom s najmenšou absolútnou hodnotou a ako posledné bude číslo s najväčšou absolútnou hodnotou. Poznamenajme však, že táto rada nemusí byť vždy najlepšia. Čísla $x_1, \ldots, x_n$ môžu mať rôzne znamienka a jednotlivé chyby sa môžu navzájom kompenzovať.

2.4 Dôsledky počítačovej aritmetiky

Detailné poznanie počítačovej aritmetiky nám môže pomôcť vyhnúť sa chybám pri programovaní niektorých logických a aritmetických operácií.

1. Testovanie na rovnosť. Obvykle nie je rozumné napísať podmienku typu

   `if` $(x=y)$ `then`
   alebo `if` $(g_1(x)=g_2(x))$ `then` .

   Pretože ak čísla x, y vznikli nejakými výpočtami, potom síce môže algebraicky platiť $x=y$, ale "počítačovo" to nemusí platiť. Preto je rozumnejšie napísať

   `if` $(|x-y|<\varepsilon)$ `then`
   resp. `if` $(|g_1(x)-g_2(x)|<\varepsilon)$ `then` .

2. Výsledok výpočtu $1\,000\,000 + \left(\sum\limits_{k=1}^{1\,000\,000} 1\right)$ nemusí byť 2 000 000 (ale 1 000 000) napr. v prípade $\beta=10, t=6$), ak počítame nasledujúcim spôsobom

   s=1 000 000
   pre k=1 až 1 000 000 s=s+1
   vystup s

   Ak však sčítame "od konca", t.j.

   s=0
   pre k=1 až 1 000 000 s=s+1
   s=s+1 000 000
   vystup s

   dostaneme správny výsledok.

3. Príkaz $y := \sqrt{x^2+1}-1$ nedáva pre "malé" x presné výsledky. Lepšie je napísať algebraicky ekvivalentný príkaz $y := \frac{x^2}{\sqrt{x^2+1}+1}$.

4. Výpočet $y := x - \sin x$ je pre malé x opäť nepresný (lebo $x \approx \sin x$). Presnejšie výsledky dostaneme, keď rozvinieme funkciu $x-\sin x$ do Taylorovho radu, ktorý usekneme po dostatočnom počte členov

$$x-\sin x = x - \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots\right).$$

5. Príkaz $y := \cos^2 x - \sin^2 x$ pre $x \approx \frac{\pi}{4}$ nedáva presné výsledky. Rozumnejšie je napísať $y := \cos 2x$.

6. Príkaz $y := \ln x - 1$ pre $x \approx \mathrm{e}$ nedáva presné výsledky. Presnejšie je napísať $y := \ln \frac{x}{\mathrm{e}}$.

7. Uvažujme kvadratickú rovnicu

$$ax^2 + bx + c = 0 \quad \text{kde } a \neq 0,\ b < 0.$$

Pre jej korene platí $r_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.
Ak $4ac \ll b^2$ potom $b \approx \sqrt{b^2 - 4ac}$ a výpočet koreňa $r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$ je nepresný. Preto je rozumnejšie vypočítať najprv koreň $r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$ a potom koreň r_2 zo vzťahu $r_1 r_2 = \frac{c}{a}$, t.j. $r_2 = \frac{c}{a r_1}$.

8. Uvažujme systém lineárnych algebraických rovníc

$$\begin{aligned} \varepsilon x_1 + x_2 &= 1, \\ x_1 + x_2 &= 2. \end{aligned}$$

Počítajme pomocou Gaussovej eliminačnej metódy

$$\left(\begin{array}{cc|c} \varepsilon & 1 & 1 \\ 1 & 1 & 2 \end{array} \right) \sim \left(\begin{array}{cc|c} \varepsilon & 1 & 1 \\ 0 & 1 - \frac{1}{\varepsilon} & 2 - \frac{1}{\varepsilon} \end{array} \right) \sim \left(\begin{array}{cc|c} \varepsilon & 1 & 1 \\ 0 & 1 & \frac{2 - \frac{1}{\varepsilon}}{1 - \frac{1}{\varepsilon}} \end{array} \right).$$

Dostávame teda

$$\begin{aligned} x_2 &= \frac{2 - \frac{1}{\varepsilon}}{1 - \frac{1}{\varepsilon}}, \\ x_1 &= \frac{1 - x_2}{\varepsilon}. \end{aligned}$$

Ak ε je "malé", potom $\frac{1}{\varepsilon}$ je "veľké", a teda pri výpočte v počítačovej aritmetike dostávame

$$\begin{aligned} x_2 &= \frac{2 - \frac{1}{\varepsilon}}{1 - \frac{1}{\varepsilon}} = 1, \\ x_1 &= \frac{1 - x_2}{\varepsilon} = \frac{1 - 1}{\varepsilon} = 0. \end{aligned}$$

Ak prehodíme poradie rovníc v pôvodnom systéme, potom

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ \varepsilon & 1 & 1 \end{array} \right) \sim \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 - \varepsilon & 1 - 2\varepsilon \end{array} \right) \sim \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & \frac{1 - 2\varepsilon}{1 - \varepsilon} \end{array} \right),$$

a teda

$$\begin{aligned} x_2 &= \frac{1 - 2\varepsilon}{1 - \varepsilon}, \\ x_1 &= 2 - x_1 = 2 - \frac{1 - 2\varepsilon}{1 - \varepsilon} = \frac{2 - 2\varepsilon - 1 + 2\varepsilon}{1 - \varepsilon} = \frac{1}{1 - \varepsilon}. \end{aligned}$$

Pre malé ε dostávame pri výpočte v počítačovej aritmetike

$$x_2 = 1, \qquad x_1 = 1,$$

čo je dobrá aproximácia správneho výsledku (lebo $\lim\limits_{\varepsilon \to 0} x_2 = 1, \lim\limits_{\varepsilon \to 0} x_1 = 1$).

9. Niekedy sa zavádza ako charakteristika strojovej presnosti číslo "strojové ε". Je to najmenšie číslo (počítačové), pre ktoré platí (v počítačovej aritmetike)

$$fl_t(1+\varepsilon) > 1.$$

Platí

$$\begin{aligned} \varepsilon &= \beta^{-t+1} && \text{v aritmetike s usekávaním} \\ \varepsilon &= \tfrac{1}{2}\beta^{-t+1} && \text{v aritmetike so zaokrúhľovaním.} \end{aligned}$$

10. Keďže systém $\mathcal{M}(t,\beta,L,U)$ obsahuje iba konečný počet čísel, môžeme dostať ako výsledok aritmetických operácií s číslami z množiny $\mathcal{M}(t,\beta,L,U)$ číslo, ktoré je väčšie ako najväčšie číslo z tejto množiny (t.j. číslo $(0,\beta\beta\cdots\beta)\beta^U$) tzv. *overflow* - niekedy hlási chybu, niekedy "NaN" resp. menšie číslo ako najmenšie (v abs. hodnote) číslo t.j. ako $0,1\beta^U = \beta^{U-1}$ - vtedy systém hlási chybu *underflow* príp. "NaN".

11. Najmenšie celé číslo m, pre ktoré platí

$$fl_t(m+1) \neq m+1$$

je číslo $m = \beta^t$.

12. Pri delení číslom, ktoré je "algebraicky kladné" môžeme dostať "delenie nulou" - hlási *floating zero divide*, resp. "NaN".

13. Výpočet funkčnej hodnoty. Nech $f: \mathbb{R} \to \mathbb{R}$ je spojite diferencovateľná funkcia (napr. polynóm). Keď napíšeme

$$y = f(x)$$

dopustíme sa chyby jednak tým, že x je zaokrúhlené a tým, že aritmetické operácie sa nevykonajú presne. Vypočítame vlastne hodnotu modifikovanej - počítačovej funkcie $\hat{f}$ v bode x_A namiesto $f(x_T)$. Chybu, ktorej sa dopustíme, môžeme odhadnúť nasledovne

$$f(x_T) - \hat{f}(x_A) = (f(x_T) - f(x_A)) + (f(x_A) - \hat{f}(x_A)).$$

Druhá zátvorka závisí na presnosti aritmetických operácií - tú sme už analyzovali. Prvú zátvorku môžeme pomocou vety o strednej hodnote napísať

$$f(x_T) - f(x_A) = f'(\xi)(x_T - x_A) \qquad \xi \text{ leží medzi } x_T, x_A.$$

Veľkosť chyby teda závisí na veľkosti derivácie f medzi x_T, x_A.

Príklad 2.10 Harmonický rad $\sum_{n=1}^{\infty} \frac{1}{n}$ je divergentný. Jeho čiastočné súčty sa dajú rekurzívne vypočítať podľa predpisu

$$s_1 = 1, \quad s_n = s_{n-1} + \frac{1}{n}; \quad n \geq 2.$$

Použitím počítačovej aritmetiky sa harmonický rad stane zdanlivo konvergentným. Určte minimálnu hodnotu indexu n tak, aby

$$s_{n+1} = fl_4^{round}\left(s_n + \frac{1}{n}\right) = s_n.$$

Vypočítajte približne súčet harmonického radu pri použití fl_4^{round} a

$$\lim_{n\to\infty}[s_n - \ln n] = C \approx 0,5772 \qquad \text{(Eulerova konštanta).}$$

Riešenie: Hľadáme n také, aby

$$\frac{1}{n} \leq 4,999 \cdot 10^{-4},$$

teda $n \geq \frac{10^4}{4,999} \approx 2001$. Pre odhad súčtu harmonického radu dostaneme

$$s_{2001} \approx \ln 2001 + 0,5772 \approx 7,601 + 0,5772 \approx 8,178\ .$$

Skutočne bude platiť

$$s_{2002} = fl_4^{round}\left(s_{2001} + \tfrac{1}{2002}\right) = s_{2001} \approx 8,178\ .$$

□

Cvičenie 2.1 Nech $a = 10^6 + 2, b = 10^6 - 1$.

(a) Vypočítajte na 10 desatinných miest

(i) $((a+b)(a-b))^2$,

(ii) $(a^2 - b^2)^2$.

(b) Aké veľké sú relatívne chyby v prípade (a)?

Cvičenie 2.2 Navrhnite spôsob výpočtu s malou zaokrúhľovacou chybou pre

$$f(x) = \frac{x\sin x}{1-\cos x}.$$

pričom $|x| \ll 1$.

Cvičenie 2.3 Nech $x_0 > -1$. Postupnosť x_n pre $n \geq 0$ je definovaná vzťahom

$$x_{n+1} = 2^{n+1}\left[\sqrt{1 + 2^{-n}x_n} - 1\right].$$

Preformulujte rekurziu tak, aby jej výpočet bol čo najpresnejší.

Kapitola 3

Rekurentné relácie

Budeme sa zaoberať výpočtom členov postupností daných rekurentnými reláciami[1]. Najvšeobecnejšia definícia rekurentnej relácie (explicitnej, nelineárnej p-teho rádu) je

$$y_{n+p} = F_{n+p}(y_n, y_{n+1}, \ldots, y_{n+p-1}), \qquad n \in \mathbb{Z}, \tag{3.1}$$

kde $F_k : \mathbb{R}^p \to \mathbb{R}$ sú dané funkcie, p je dané kladné číslo.

Lineárna rekurentná relácia p-teho rádu má tvar

$$y_{n+p} = \tilde{b}_{p-1,n+p} y_{n+p-1} + \tilde{b}_{p-2,n+p} y_{n+p-2} + \ldots + \tilde{b}_{0,n+p} y_n + a_{n+p}, \tag{3.2}$$

kde $\tilde{b}_{i,n+p}$; $i = 0, \ldots, p-1$; a_{n+p}, $n = 0, 1, \ldots$ sú dané čísla, pričom $\tilde{b}_{0,n+p} \neq 0$.
Ak $a_k = 0$ pre $k \in \mathbb{Z}$, potom hovoríme, že relácia (3.2) je *homogénna*.
Ak koeficienty $\tilde{b}_{i,n+p}$ nezávisia na n potom môžeme písať

$$y_{n+p} = \tilde{b}_{p-1} y_{n+p-1} + \tilde{b}_{p-2} y_{n+p-2} + \ldots + \tilde{b}_0 y_n + a_{n+p}. \tag{3.3}$$

Hovoríme, že (3.3) je *lineárna rekurentná* relácia p-teho rádu s konštantými koeficientami (nehomogénna). Ak $a_k = 0$ pre $k \in \mathbb{Z}$, potom je homogénna.

Príklad 3.1 $y_{n+1} = y_n + 1$, $n \in \mathbb{Z}$ je nehomogénna lineárna rekurentná relácia prvého rádu s konštantnými koeficientami. □

Príklad 3.2 $y_{n+2} = y_{n+1} + y_n$, $n \in \mathbb{Z}$ je homogénna lineárna rekurentná relácia druhého rádu s konštantnými koeficientami. □

Príklad 3.3 $y_{n+3} = n y_{n+2} y_{n+1} - y_n + 1$, $n \in \mathbb{Z}$ je nehomogénna nelineárna rekurentná relácia tretieho rádu s premennými koeficientami. □

Ďalej sa budeme zaoberať lineárnou rekurentnou reláciou (LRR) prvého rádu a druhého rádu s konštantnými koeficientami.

3.1 Lineárna rekurentná relácia 1. rádu

Lineárna rekurentná relácia 1. rádu má tvar

$$y_n = b_n y_{n-1} + a_n, \qquad n \in \mathbb{Z}. \tag{3.4}$$

O riešení úlohy (3.4) v explicitnom tvare hovorí nasledujúca veta.

[1] V literatúre sa niekedy používa pomenovanie diferenčné rovnice.

Veta 3.1 *Nech $b_n \neq 0$ pre $n \in \mathbb{Z}$, nech $c \in \mathbb{C}$, $n_0 \in \mathbb{Z}$. Potom postupnosť*

$$y_n = \left(\frac{c}{b_{n_0}} + \frac{a_{n_0+1}}{\prod\limits_{i=n_0}^{n_0+1} b_i} + \frac{a_{n_0+2}}{\prod\limits_{i=n_0}^{n_0+2} b_i} + \ldots + \frac{a_n}{\prod\limits_{i=n_0}^{n} b_i} \right) \prod_{i=n_0}^{n} b_i \tag{3.5}$$

spĺňa

$$\begin{aligned} y_n &= b_n y_{n-1} + a_n, \qquad n > n_0, \\ y_{n_0} &= c. \end{aligned} \tag{3.6}$$

DÔKAZ: Použijeme matematickú indukciu.

Pre $n = n_0$ je tvrdenie triviálne.

Nech je tvrdenie pravdivé pre n. Dokážeme jeho platnosť pre $n+1$:

$$\begin{aligned} y_{n+1} &= b_{n+1} y_n + a_{n+1} \\ &= b_{n+1} \left(\frac{c}{b_{n_0}} + \frac{a_{n_0+1}}{\prod\limits_{i=n_0}^{n_0+1} b_i} + \ldots + \frac{a_n}{\prod\limits_{i=n_0}^{n} b_i} \right) \prod_{i=n_0}^{n} b_i + a_{n+1} \\ &= \left(\frac{c}{b_{n_0}} + \frac{a_{n_0+1}}{\prod\limits_{i=n_0}^{n_0+1} b_i} + \ldots + \frac{a_n}{\prod\limits_{i=n_0}^{n} b_i} + \frac{a_{n+1}}{\prod\limits_{i=n_0}^{n+1} b_i} \right) \prod_{i=n_0}^{n+1} b_i. \end{aligned}$$

□

Príklad 3.4 Lineárna rekurentná relácia $y_n = q y_{n-1}$ má riešenie $y_n = c q^n$. □

Príklad 3.5 Nech na začiatku sporenia máme v banke y_0 korún. Nech ročný úrok je $r\%$, pričom úroky sa pripisujú mesačne a nech sa mesačne platí d korún poplatkov. Označme y_1 sumu, ktorú budeme mať v banke na konci 1. mesiaca, potom na konci n-tého mesiaca bude teda na účte

$$y_n = y_{n-1} + \frac{r}{12} y_{n-1} - d \quad \text{korún.}$$

Ak označíme $a = 1 + \frac{r}{12}, b = -d$ dostávame teda lineárnu rekurentnú reláciu 1. rádu s konštantnými koeficientami

$$y_n = a y_{n-1} + b. \tag{3.7}$$

Riešením je

$$\begin{aligned} y_n &= a y_{n-1} + b && = a(a y_{n-2} + b) + b \\ &= a^2 y_{n-2} + ab + b && = a^2(a y_{n-3} + b) + ab + b \\ &= a^3 y_{n-3} + a^2 b + ab + b && = \ldots \\ &= a^n y_0 + b(a^{n-1} + a^{n-2} + \ldots + a + 1) && = a^n y_0 + b\tfrac{1-a^n}{1-a}, \end{aligned}$$

t.j.

$$y_n = a^n y_0 + b \frac{1 - a^n}{1 - a} \tag{3.8}$$

(za predpokladu, že $a \neq 1$).

Pozrime teraz na stabilitu relácie (3.7) v počítačovej realizácii. Kvôli jednoduchosti predpokladajme, že čísla a, b patria do množiny počítačových čísel, Y_0 je zaokrúhlená hodnota hodnoty y_0. V skutočnosti budeme počítať namiesto

$$y_{n+1} = ay_n + b$$

hodnoty

$$Y_{n+1} = aY_n + b + R_{n+1}, \qquad n = 0, 1, 2, \dots \ .$$

Člen R_{n+1} reprezentuje zaokrúhľovacie chyby pri výpočte. Skúmajme teraz akumulovanú chybu, t.j. $Y_{n+1} - y_{n+1}$. Zrejme platí

$$\begin{aligned}
Y_{n+1} - y_{n+1} &= aY_n + b + R_{n+1} - (ay_n + b) \\
&= a(Y_n - y_n) + R_{n+1} \\
&= a\left(a\left(Y_{n-1} - y_{n-1}\right) + R_n\right) + R_{n+1} \\
&= a^2\left(Y_{n-1} - y_{n-1}\right) + aR_n + R_{n+1} \\
&= a^2\left(a\left(Y_{n-2} - y_{n-2}\right) + R_{n-1}\right) + aR_n + R_{n+1} \\
&= a^3\left(Y_{n-2} - y_{n-2}\right) + a^2R_{n-1} + aR_n + R_{n+1} \\
&= \dots \\
&= a^{n+1}\left(Y_0 - y_0\right) + \dots + a^2R_{n-1} + aR_n + R_{n+1} \ .
\end{aligned}$$

Teda

Ak $|a| < 1$ a $|R_n| \leq R \quad$ pre $n = 1, 2, \dots$, potom $|Y_n - y_n| \leq c \qquad \forall n$.

Ak $|a| > 1$, potom $|Y_n - y_n| \to \infty \quad$ pre $n \to \infty$.

Ak $|a| = 1$ a ak $R_n \geq R_0 > 0 \quad \forall n$, potom $|Y_n - y_n| \to \infty$ pre $n \to \infty$.

V našom prípade $|a| = 1 + \frac{r}{12} > 1$, a teda chyba rastie do nekonečna. □

Príklad 3.6 V sklenenej nádobe je tekutina, ktorá má v čase t_0 teplotu y_0. Nech okolité prostredie má teplotu $\bar{y}$ a nech c je tepelná vodivosť skla, z ktorého je nádoba. Označme y_n teplotu tekutiny v čase $t_n = t_0 + nh$, kde $h > 0$ je malé. Podľa Newtonovho zákona platí

$$y_{n+1} - y_n = ch(\bar{y} - y_n,) \qquad n = 0, 1, 2, \dots,$$

t.j.

$$y_{n+1} = (1 - ch)y_n + ch\bar{y}, \qquad n = 0, 1, 2, \dots,$$

alebo inak zapísané

$$y_{n+1} = ay_n + b,$$

kde $a = 1 - ch$, $b = ch\bar{y}$.

Fyzikálne zrejme platí, že "v nekonečnom čase" bude teplota tekutiny rovnaká ako v okolitom prostredí. Skúmajme, či to platí aj pre našu postupnosť y_n. Zrejme

$$\begin{aligned}
y_{n+1} &= ay_n + b = a\left(ay_{n-1} + b\right) + b = a^2y_{n-1} + ab + b = \\
&= \dots = a^{n+1}y_0 + b\left(a^n + a^{n-1} + \dots + a + 1\right) = \\
&= a^{n+1}y_0 + b\frac{1 - a^{n+1}}{1 - a}.
\end{aligned}$$

To znamená, že ak $|a| < 1$, potom $y_n \to \frac{b}{1-a} = \frac{ch\bar{y}}{1-(1-ch)} = \bar{y}$, čo je v súlade s fyzikálnou realitou.

Ak $|a| \geq 1$, potom $y_n \to \infty$, čo je v rozpore s fyzikálnou realitou.

Keďže $c > 0, h > 0$ podmienka $|a| < 1$ znamená $|1 - ch| < 1$, t.j. $ch < 2$. Časový krok h je teda potrebné voliť tak, aby $h < \frac{2}{c}$. □

Poznámka 3.1 *Ak poznáme hodnoty y_0, y_1, potom môžeme konštantu c vypočítať z rovnosti* $y_1 - y_0 = ch(\bar{y} - y_0)$. □

3.2 Lineárna rekurentná relácia 2. rádu s konštantnými koeficientami

Uvažujme najprv homogénnu lineárnu rekurentnú reláciu

$$y_{n+2} + b_1 y_{n+1} + b_2 y_n = 0. \tag{3.9}$$

Hľadajme riešenie (3.9) v tvare

$$y_n = z^n. \tag{3.10}$$

Hľadáme teda také $z \neq 0$, aby platilo

$$z^{n+2} + b_1 z^{n+1} + b_2 z^n = 0, \qquad \text{t.j.} \qquad z^n(z^2 + b_1 z + b_2) = 0.$$

Ak je číslo z koreňom polynómu

$$p(z) = z^2 + b_1 z + b_2, \tag{3.11}$$

potom postupnosť (3.10) je riešením (3.9). Teraz sú dve možnosti:

i) Ak polynóm p má dva rôzne korene z_1, z_2, potom $\{z_1^n\}$, $\{z_2^n\}$ sú dve riešenia (3.9).

ii) Ak polynóm p má dvojnásobný koreň z, potom $\{z^n\}$ je riešením (3.9). V tomto prípade je druhé riešenie rovnice (3.9) postupnosť

$$y_n = nz^{n-1}. \tag{3.12}$$

Platí totiž

$$(n+2)z^{n+1} + b_1(n+1)z^n + b_2 n z^{n-1} = nz^{n-1}(z^2 + b_1 z + b_2) + z^n(2z + b_1) = 0,$$

lebo z je koreňom p aj p'.

Platí teda nasledujúce tvrdenie.

Veta 3.2 *Nech $b_1, b_2 \in \mathbb{R}$. Potom existujú dve lineárne nezávislé postupnosti spĺňajúce (3.9).*

DÔKAZ: Stačí dokázať lineárnu nezávislosť, lebo existenciu sme už ukázali.

Prípad *i)*. Nech $\{z_1^n\}, \{z_2^n\}$ sú lineárne závislé. Potom existuje α tak, že $z_1^n = \alpha z_2^n$, pre $n \in \mathbb{Z}$, t.j. $z_1^0 = \alpha z_2^0, z_1^1 = \alpha z_2^1$, čo je spor s tým, že $z_1 \neq z_2$.

Prípad *ii)*. Podobne ako v prípade *i)*, ak $\{z^n\}, \{nz^{n-1}\}$ sú lineárne závislé postupnosti, potom $z^0 = 0, z^1 = z^0 = 1$, čo je spor.

□

Dôsledok 3.1 *Množina všetkých postupností spĺňajúcich (3.9) je lineárny vektorový priestor s dimenziou aspoň 2.* □

Veta 3.3 *Dimenzia lineárneho vektorového priestoru všetkých postupností spĺňajúcich (3.9) je 2.*

DÔKAZ: Stačí ukázať, že dimenzia je najviac 2.

Nech X, Y sú dve lineárne nezávislé postupnosti spĺňajúce (3.9). Ukážeme, že ak Z je postupnosť spĺňajúca (3.9), potom existujú d_1, d_2 také, že $Z = d_1X + d_2Y$. Definujme d_1, d_2 takto: Vektor $(d_1, d_2) \in \mathbb{R}^2$ je riešením systému dvoch lineárnych algebraických rovníc o dvoch neznámych

$$d_1x_0 + d_2y_0 = z_0,$$
$$d_1x_1 + d_2y_1 = z_1,$$

t.j.

$$\begin{pmatrix} x_0 & y_0 \\ x_1 & y_1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} z_0 \\ z_1 \end{pmatrix}. \tag{3.13}$$

Z konštrukcie bázy X, Y vyplýva, že problém má práve jedno riešenie. Definujme teraz

$$U = d_1X + d_2Y.$$

Ukážeme, že platí $U = Z$. Zrejme z (3.13) vyplýva, že $z_0 = u_0, z_1 = u_1$. Preto zo vzťahov (3.13) a (3.9) vyplýva

$$z_2 = -b_1z_1 - b_2z_0 = -b_1u_1 - b_2u_0 = u_2,$$

atď. matematickou indukciou. Podobne $z_{-1} = u_{-1}$. □

Veta 3.4 *Nech $b_1, b_2, A, B \in \mathbb{R}, n_0 \neq n_1 \in \mathbb{Z}$. Potom existuje práve jedna postupnosť $\{y_n\}$ spĺňajúca (3.9) a navyše $y_{n_0} = A, y_{n_1} = B$.*

DÔKAZ: Každá postupnosť Z spĺňajúca (3.13) sa dá písať v tvare $Z = d_1X + d_2Y$, kde X, Y je báza skonštruovaná v texte pred vetou 3.2. Koeficienty d_1, d_2 zvolíme tak, aby platilo

$$d_1x_{n_0} + d_2y_{n_0} = A,$$
$$d_1x_{n_1} + d_2y_{n_1} = B,$$

t.j. v prípade *i)*

$$\begin{pmatrix} z_1^{n_0} & z_2^{n_0} \\ z_1^{n_1} & z_2^{n_1} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} \tag{3.14}$$

a v prípade *ii)*

$$\begin{pmatrix} z_0^n & n_0z^{n_0-1} \\ z_1^n & n_1z^{n_1-1} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix}. \tag{3.15}$$

Matice systémov (3.14), (3.15) sú zrejme regulárne, čím je veta dokázaná. □

Uvažujme teraz nehomogénnu rovnicu

$$y_{n+2} + b_1y_{n+1} + b_2y_n = a_{n+2}. \tag{3.16}$$

Veta 3.5 (princíp superpozície)

i) Nech V je nejaká pevne zvolená postupnosť spĺňajúca (3.16). Potom každá postupnosť Y spĺňajúca (3.16) sa dá napísať v tvare

$$Y = U + V,$$

kde U je vhodná postupnosť spĺňajúca (3.9).

ii) Naopak nech V spĺňa (3.16) a U spĺňa (3.9), potom

$$Y = U + V$$

spĺňa (3.16).

DÔKAZ:

i) Nech Y spĺňa (3.16). Dosadením dostávame, že $U = Y - V$ spĺňa (3.9).

ii) Táto časť sa dokáže dosadením. □

Každú postupnosť spĺňajúcu (3.16) môžeme teda dostať, ak vieme riešiť (3.9) a ak poznáme aspoň jednu postupnosť spĺňajúcu (3.16). Teraz sa zameriame na hľadanie nejakej postupnosti spĺňajúcej (3.16). Najprv budeme uvažovať dva špeciálne prípady pravej strany a_n.

Prípad 1: Nech a_n je polynóm. Potom hľadáme riešenie v tvare polynómu.

Príklad 3.7

$$y_{n+2} - y_{n+1} - 2y_n = 2n^2 + 2, \tag{3.17}$$

t.j. $a_{n+2} = 2n^2 + 2$. Hľadáme postupnosť y_n v tvare $y_n = An^2 + Bn + C$. Dosadením do (3.17) dostávame $A = -1, B = -1, C = -3$, t.j.

$$y_n = -n^2 - n - 3.$$ □

Prípad 2: Nech $a_n = cq^n$. Riešenie hľadáme v tvare

$$y_n = dq^n.$$

Po dosadení dostávame

$$dq^n \left(q^2 + b_1 q + b_2\right) = cq^{n+2}.$$

Označme $p(q) = q^2 + b_1 q + b_2$.

Ak $p(q) \neq 0$, potom $d = cq^2/p(q)$.

Ak $p(q) = 0$, potom hľadáme riešenie v tvare $y_n = dnq^{n-1}$. Po dosadení dostávame

$$dnq^{n-1}p(q) + dq^n p'(q) = cq^{n+2},$$

a teda

ak $p'(q) \neq 0$, potom $d = cq^2/p'(q)$.

ak $p'(q) = 0$, potom hľadáme riešenie v tvare $y_n = n^2 a q^{n-2}$. Po dosadení dostávame

$$n^2 d q^{n-2} p(q) + 2nd q^{n-1} p'(q) + q^{n-1} d(4q + b_1) = c q^{n+2},$$

t.j. $d = cq^3/(4q + b_1)$. □

V dvoch špeciálnych prípadoch teda vieme nájsť aspoň jednu postupnosť spĺňajúcu (3.16). Všeobecný prípad rieši nasledujúca veta.

Veta 3.6 *Nech $\{a_n\}$ je daná postupnosť (komplexných) čísel, nech $n_0 \in \mathbb{Z}, a_{n_0} \neq 0$. Nech z je riešenie rovnice*

$$z_{n+2} + b_1 z_{n+1} + b_2 z_n = 0$$

spĺňajúce $z_0 = 1, z_{-1} = 0$. *Potom postupnosť $Y = \{y_n\}_{n=n_0-1}^{\infty}$ definovaná pomocou*

$$y_{n_0+p} = \sum_{j=0}^{p} z_{p-j} a_{n_0+j} \tag{3.18}$$

spĺňa

$$y_{n+2} + b_1 y_{n+1} + b_2 y_n = a_{n+2}, \tag{3.19}$$

$$y_{n_0-1} = 0, \ y_{n_0} = a_{n_0}. \tag{3.20}$$

DÔKAZ: Stačí dosadiť (3.18) do (3.19), (3.20). □

Príklad 3.8 Nech

$$y_{n+2} - 5y_{n+1} + 6y_n = 0. \tag{3.21}$$

Riešenie hľadáme v tvare $y_n = az^n$. Dostaneme

$$z^2 - 5z + 6 = (z-2)(z-3) = 0,$$
$$z_1 = 2, \ z_2 = 3.$$

Všeobecné riešenie je teda

$$y_n = d_1 2^n + d_2 3^n.$$

Hľadajme teraz riešenie (3.21) spĺňajúce

$$y_0 = 1, \ y_1 = 1\ . \tag{3.22}$$

Teda

$$\begin{aligned} d_1 2^0 &+ d_2 3^0 &= 1 \\ d_1 2 &+ d_2 3 &= 1. \end{aligned}$$

Odtiaľ vyplýva $d_1 = 2$, $d_2 = -1$. Riešenie úlohy (3.21), (3.22) je teda postupnosť $y_n = 2^{n+1} - 3^n$. □

Príklad 3.9 Určte y_n tak, aby

$$y_{n+2} - 2y_{n+1} + y_n = 0, \qquad y_0 = 1, \ y_1 = 0.$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Odtiaľ

$$\begin{aligned} z^2 - 2z + 1 &= 0 \\ (z-1)^2 &= 0 \qquad \text{dvojnásobný koreň.} \end{aligned}$$

Všeobecné riešenie je teda

$$y_n = d_1 + d_2 n.$$

Koeficienty d_1, d_2 majú spĺňať

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

t.j. $d_1 = 1,\ d_2 = -1$. Riešenie je

$$y_n = 1 - n.$$

□

Príklad 3.10 Určte y_n tak, aby

$$y_{n+2} + y_n = 0. \tag{3.23}$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Platí

$$\begin{aligned} z^2 + 1 &= 0, \\ z_1 = i,\ z_2 &= -i. \end{aligned}$$

Všeobecné riešenie je

$$d_1 i^n + d_2(-i)^n.$$

Reálnu bázu priestoru postupností spĺňajúcich (3.23) dostaneme pomocou partikulárnych riešení takto: $y_n = \dfrac{i^n + (-i)^n}{2}$ a $y_n = \dfrac{i^n - (-i)^n}{2i}$, čiže

$$\begin{array}{lllll} 1, & 0, & -1, & 0, & \ldots \\ 0, & 1, & 0, & -1, & \ldots . \end{array}$$

□

Príklad 3.11 Vyriešte

$$y_{n+2} = y_{n+1} + y_n.$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Teda

$$\begin{aligned} z^2 - z - 1 &= 0, \\ z_{1,2} &= \frac{1}{2} \pm \frac{\sqrt{5}}{2}. \end{aligned}$$

Všeobecné riešenie je

$$y_n = d_1 \left(\frac{1}{2} + \frac{\sqrt{5}}{2} \right)^n + d_2 \left(\frac{1}{2} - \frac{\sqrt{5}}{2} \right)^n.$$

□

Príklad 3.12 Nájdite y_n tak, aby

$$y_{n+2} - 6y_{n+1} + 9y_n = 0, \qquad y_0 = 1,\ y_1 = 2.$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Potom

$$z^2 - 6z + 9 = 0,$$

teda

$$(z-3)^2 = 0.$$

Všeobecné riešenie má tvar

$$y_n = d_1 3^n + d_2 n 3^{n-1}.$$

Koeficienty d_1, d_2 majú spĺňať

$$\begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

t.j. $d_1 = 1,\ d_2 = -1$. Riešenie je

$$y_n = 3^n - 3^{n-1} n.$$

Príklad 3.13 Vyriešte

$$y_{n+2} = \frac{10}{3} y_{n+1} - y_n, \qquad y_0 = A,\ y_1 = B.$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Odtiaľ

$$z^2 - \frac{10}{3} z + 1 = 0,$$

teda

$$z_1 = 3,\ z_2 = \frac{1}{3}.$$

Všeobecné riešenie je

$$y_n = d_1 3^n + d_2 3^{-n}.$$

Koeficienty d_1, d_2 majú spĺňať

$$\begin{pmatrix} 1 & 1 \\ 3 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix},$$

t.j. $d_1 = \frac{3B-A}{8},\ d_2 = \frac{9A-3B}{8}$. Riešenie je

$$y_n = \frac{3B-A}{8} 3^n + \frac{3A-B}{8} 3^{1-n}.$$

Ak $A \neq 3B$, potom je riešenie nestabilné, t.j. malá zmena implikuje veľkú chybu. □

Príklad 3.14 Určte y_n tak, aby

$$y_{n+2} = -y_{n+1} + y_n, \quad y_0 = 1,\ y_1 = \frac{\sqrt{5}-1}{2}.$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Teda

$$z^2 + z - 1 = 0,$$

teda

$$z_1 = -\frac{1}{2} + \frac{\sqrt{5}}{2}, \quad z_2 = -\frac{1}{2} - \frac{\sqrt{5}}{2}.$$

Platí

$$\frac{1}{2} < z_1 < 1, \qquad -\frac{3}{2} < z_2 < -1.$$

Všeobecné riešenie je

$$y_n = d_1 \left(\frac{\sqrt{5}-1}{2} \right)^n + d_2 \left(-\frac{\sqrt{5}+1}{2} \right)^n .$$

Ak hľadáme riešenie spĺňajúce začiatočné podmienky, tak $d_1 = 1,\ d_2 = 0$.
Ak $d_2 = \varepsilon \neq 0$ prevláda z_2^n, t.j. nestabilita. □

Príklad 3.15 Vyriešte nehomogénnu úlohu

$$y_{n+2} - y_{n+1} - 2y_n = 2n^2 + 2, \tag{3.24}$$

$$y_0 = 0,\ y_1 = 1. \tag{3.25}$$

Riešenie: Jedno riešenie nehomogénnej rovnice je $y_n = -n^2 - n - 3$ (prípad 1). Všeobecné riešenie príslušnej homogénnej úlohy je

$$d_1 2^n + d_2(-1)^n,$$

lebo

$$z^2 - z - 2 = 0$$

má riešenie

$$z_1 = 2, \quad z_2 = -1.$$

Všeobecné riešenie (3.24) je

$$d_1 2^n + d_2(-1)^n - n^2 - n - 3.$$

Riešenie (3.24), (3.25) je

$$y_n = 3 \cdot 2^n - n^2 - n - 3.$$ □

Príklad 3.16 Nájdite y_n tak, aby

$$y_{n+2} - 2y_{n+1} + y_n = 1, \qquad y_0 = 5,\ y_1 = 8.$$

Riešenie: Riešenie hľadáme v tvare $y_n = az^n$. Čiže

$$z^2 - 2z + 1 = (z-1)^2 = 0.$$

Všeobecné riešenie homogénneho problému je $z_n = d_1 + d_2 n$. Riešenie homogénneho problému spĺňajúce $z_{-1} = 0,\ z_0 = 1$ je $z_n = 1 + n$.

Aplikujme vetu 3.6 pre $n_0 = 1 = a_n,\ \forall n;\ p = n - 1$. Platí

$$y_n = y_{1+n-1} = \sum_{j=0}^{n-1} z_{p-j}.1 = \sum_{j=0}^{n-1} [(1+n-1) - j] = \sum_{j=0}^{n-1} (n-j) = \frac{n(n+1)}{2} .$$

Potrebujeme ešte napasovať začiatočné podmienky. Riešenie hľadáme v tvare

$$y_n = d_1 + d_2 n + \frac{1}{2}n(n+1).$$

Pre d_1, d_2 dostávame rovnice

$$\begin{aligned} d_1 \quad &= 5 \\ d_1 + d_2 + 1 \quad &= 8 \Rightarrow d_2 = 2, \end{aligned}$$

a teda

$$y_n = 5 + 2n + \frac{1}{2}n(n+1)\ .$$

□

3.3 Gronwallova lema

V tomto odseku sa budeme zaoberať spojitou i diskrétnou verziou Gronwallovej lemy. Spojitá verzia hovorí o tom ako z určitej relácie (nerovnosti) medzi funkciou a jej deriváciou sa dá dostať odhad pre samotnú funkciu. V diskrétnej verzii si odvodíme odhad členov číselnej postupnosti spĺňajúcej rekurentné nerovnosti. S aplikáciami oboch verzií Gronwallovej lemy sa môžme stretnúť pri získavaní apriórnych odhadov riešenia obyčajných diferenciálnych rovníc, parabolických parciálnych diferenciálnych rovníc atď.

Lema 3.1 (Gronwallova) *Nech $r(t), h(t), y(t)$ sú spojité reálne funkcie definované na intervale $[a, b]$, pričom $r(t), h(t) \geq 0$.*

(i) Ak platí

$$y(t) \leq h(t) + \int_a^t r(s)y(s)\ ds \qquad \text{pre} \qquad a \leq t \leq b,$$

potom

$$y(t) \leq h(t) + \int_a^t h(s)r(s)\exp\left(\int_s^t r(\tau)d\tau\right)\ ds$$

je splnené pre všetky $t \in [a, b]$.

(ii) Ak v prípade (i) navyše platí $r(s) = C$ a funkcia h je neklesajúca, tak

$$y(t) \leq h(t)e^{C(t-a)} \qquad \text{pre} \qquad a \leq t \leq b.$$

Dôkaz: (*i*) Označme $z(t) = \displaystyle\int_a^t r(s)y(s)ds$. Potom existuje z' a platí

$$z'(t) - r(t)z(t) \leq h(t)r(t).$$

Označme teraz $w(t) = z(t)\exp\left(-\displaystyle\int_a^t r(s)ds\right)$. Z poslednej nerovnosti vyplýva

$$w'(t) \leq h(t)r(t)\exp\left(-\int_a^t r(s)ds\right).$$

Pretože $w(a) = 0$, integrovaním dostaneme

$$w(t) \leq \int_a^t h(s)r(s)\exp\left(-\int_a^s r(\tau)d\tau\right)\ ds,$$

teda

$$z(t) \leq \int_a^t h(s)r(s)\exp\left(\int_s^t r(\tau)d\tau\right) ds.$$

Pomocou $y(t) \leq h(t) + z(t)$ dostaneme napokon požadované tvrdenie.

(*ii*) Jednoduchým výpočtom máme

$$\begin{aligned} y(t) &\leq h(t) + \int_a^t h(s)r(s)\exp\left(\int_s^t r(\tau)d\tau\right) ds \\ &= h(t) + C\int_a^t h(s)e^{C(t-s)}ds \\ &\leq h(t) + h(t)C\int_a^t e^{C(t-s)}ds \\ &= h(t)e^{C(t-a)}. \end{aligned}$$

□

Lema 3.2 (diskrétna verzia) *Nech $\{A_i\}, \{a_i\}$ sú postupnosti nezáporných reálnych čísel a $q \geq 0$. Ak pre $i \in \mathbb{N}$ platí*

$$a_i \leq A_i + \sum_{j=1}^{i-1} a_j q, \tag{3.26}$$

potom

$$a_i \leq A_i + e^{qi}\sum_{j=1}^{i-1} A_j q, \qquad i \in \mathbb{N}.$$

DÔKAZ: Pre $i \geq 2$ definujme

$$S(n,i) = \begin{cases} \displaystyle\sum_{j=1}^{i-1}\sum_{k_1=1}^{j-1}\sum_{k_2=1}^{k_1-1}\cdots\sum_{k_n=1}^{k_{n-1}-1} A_{k_n} & \text{pre } 0 < n \leq i-2, \\ \displaystyle\sum_{j=1}^{i-1} A_j & \text{pre } n = 0, \\ 0 & \text{pre } n > i-2. \end{cases}$$

Ak použijeme rekurziu $(i-1)$-krát v (3.26), dostaneme

$$a_i \leq A_i + q\sum_{k=0}^{i-1} q^k S(k,i).$$

Matematickou indukciou vzhľadom na n a použitím faktu, že $\binom{x}{y} = 0$ pre $x < y$, sa dá dokázať

$$S(n,i) = \sum_{j=1}^{i-1} A_j \binom{i-1-j}{n}, \qquad n \geq 0. \tag{3.27}$$

Takže môžme písať

$$
\begin{aligned}
a_i & \leq A_i + q\sum_{j=1}^{i-1} A_j \left[1 + \sum_{k=1}^{i-j-1} q^k \binom{i-1-j}{k}\right] \\
& = A_i + q\sum_{j=1}^{i-1} A_j \left[1 + \sum_{k=1}^{i-j-1} \frac{q^k (i-j)^k}{k!} \prod_{l=1}^{k} \left(1 - \frac{l}{i-j}\right)\right] \\
& \leq A_i + q\sum_{j=1}^{i-1} A_j e^{q(i-j)} \\
& \leq A_i + e^{qi} \sum_{j=1}^{i-1} A_j q.
\end{aligned}
$$

□

Príklad 3.17 Ukážte, že riešenie $u(t)$ parabolickej diferenciálnej rovnice tvaru

$$
\begin{aligned}
\frac{du(t)}{dt} - au(t) &= f(t), \qquad t \in (0,T), \\
u(0) &= u_0
\end{aligned}
\tag{3.28}
$$

je ohraničená funkcia, t.j. $\exists C > 0 : \ |u(t)| \leq C$ pre $t \in [0,T]$.

Riešenie: Integrovaním diferenciálnej rovnice (3.28) a použitím Newtonovho-Leibnitzovho vzorca dostaneme

$$u(t) - u_0 = \int_0^t [au(s) + f(s)]ds \ .$$

Označme $h(t) := \int_0^t |f(s)|\, ds + |u_0|$. Potom

$$|u(t)| \leq h(t) + |a| \int_0^t |u(s)|\, ds \ .$$

Použitím spojitej verzie Gronwallovej lemy 3.1 dostaneme pre riešenie u odhad

$$|u(t)| \leq h(t) + |a| \int_0^t h(s) \mathrm{e}^{t-s} ds \ .$$

Za predpokladu, že $f \in C([0,T])$ existuje konštanta $M > 0$ taká, že

$$\max_{0 \leq t \leq T} h(t) \leq M \ .$$

Potom

$$
\begin{aligned}
|u(t)| & \leq M + |a|\, M \int_0^t \mathrm{e}^{t-s} ds \\
& \leq M + |a|\, M \int_0^t \mathrm{e}^{T} ds \\
& \leq M + |a|\, M T \mathrm{e}^{T} =: C \ .
\end{aligned}
$$

□

Cvičenie 3.1 Nech $\boldsymbol{A} = (a_{i,j})_{n\times n}$ je daná štvorcová matica. Jej determinant sa dá rekurentne vypočítať pomocou rozvoja podľa niektorého riadka. Analyzujte výpočtovú zložitosť tohto rekurentného výpočtu, t.j. vyjadrite počet operácií (sčitovaní, násobení) ako funkciu n.

Cvičenie 3.2 Dokážte (3.27) pomocou matematickej indukcie vzhľadom na n.

Cvičenie 3.3 V rovine je daných n priamok vo všeobecnej pozícii, t.j. žiadne dve nie sú paralelné a žiadne tri sa nepretínajú v jednom bode. Nech a_n je počet podoblastí, na ktoré je rovina týmito priamkami rozdelená. Zrejme $a_0 = 1$, $a_1 = 2$, $a_2 = 4$, $a_3 = 7, \ldots$. Nájdite rekurentnú reláciu pre a_n a pomocou nej odvoďte explicitnú formulku pre a_n.

Cvičenie 3.4 Lucasova postupnosť je daná rekurentnou reláciou

$$L_n = L_{n-1} + L_{n-2}, \qquad n \geq 2$$

so začiatočnými podmienkami $L_0 = 2, L_1 = 1$. Odvoďte explicitnú formulku pre L_n pre $n \geq 0$.

Cvičenie 3.5 Daná je rekurentná relácia

$$a_n = 2a_{n-1} + 1, \qquad n \geq 1$$

so začiatočnou podmienkou $a_0 = 0$. Nájdite explicitný vzorec pre a_n ak $n \geq 0$.

Kapitola 4

Aproximácia funkcií

Predstavme si, že našou úlohou je opísať' rozloženie pôdneho znečistenia danou chemikáliou. K dispozícii máme meracie prístroje. Jednotlivými vrtmi odoberieme vzorky pôdy, ktoré potom podrobíme analýze. Problém spočíva v tom, že nie je možné aby sme takto "zmapovali" celú oblasť' (na to nikdy nie je dosť' peňazí ani času). Takže nakoniec sa musíme uspokojiť' len s konečným počtom meraní. Našou úlohou je však získať' dostatočne presný opis znečistenia celého územia. Musíme teda nejakým spôsobom "predĺžiť' "namerané hodnoty na celú oblasť'. Samozrejme toto predĺženie, pri zachovaní nameraných hodnôt, sa dá urobiť' mnohými spôsobmi (lineárnou aproximáciou, polynómami, exponenciálnou funkciou,...). Vhodná voľba závisí od konkrétnej situácie i fyzikálnej intuície. V tejto kapitole si rozoberieme niektoré možnosti.

4.1 Interpolácia polynómom

Uvažujme nasledujúcu úlohu: Nech sú dané čísla $x_i, y_i \in \mathbb{R}$ pre $i = 0, 1, 2, \ldots, n$, pričom $x_i \neq x_j$ pre $i \neq j$. Hľadáme polynóm p stupňa najviac n, pre ktorý platí

$$p(x_i) = y_i, \qquad i = 0, 1, 2, \ldots, n. \tag{4.1}$$

Motivácia:

a) y_i sú namerané hodnoty nejakej veličiny závislej na nezávisle premennej x, chceme cez ne preložiť' funkciu

b) $y_i = f(x_i)$, chceme jednoduchšie, resp. približné vyjadrenie f.

Označenie: P_n značí množinu všetkých polynómov stupňa najviac n.

Veta 4.1 *Nech $x_i, y_i \in \mathbb{R}$ pre $i = 0, 1 \ldots, n$, pričom $x_i \neq x_j$ pre $i \neq j$. Potom existuje práve jeden polynóm $p \in P_n$ tak, že platí (4.1).*

DÔKAZ:

Jednoznačnosť': Nech p_1, $p_2 \in P_n$ spĺňajú (4.1). Potom polynóm $r = p_1 - p_2$ spĺňa $r \in P_n$, $\quad r(x_i) = 0, \quad i = 0, 1, \ldots, n$, a preto $r \equiv 0$, t.j. $p_1 \equiv p_2$.

Existencia: Hľadáme polynóm p v tvare

$$p(x) = \sum_{i=0}^{n} a_i x^i.$$

Zo vzťahu (4.1) dostaneme systém $n+1$ lineárnych algebraických rovníc pre $n+1$ neznámych, ktoré môžeme zapísať takto

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ & & \vdots & & \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{4.2}$$

Z jednoznačnosti, ktorú sme práve dokázali vyplýva, že matica tohto systému má nenulový determinant a z toho dostaneme existenciu pre ľubovoľné hodnoty $y_0, y_1, \dots, y_n$. □

Poznámka 4.1 *Polynóm p z vety 4.1 sa nazýva interpolačný polynóm.*

Poznámka 4.2 *Uvedená matica sa nazýva Vandermondova matica a pre jej determinant platí* $\det = \prod_{i \neq j} (x_i - x_j)$.

Výpočet koeficientov a_i zo systému lineárnych rovníc v predchádzajúcej vete je dosť nepohodlný. Preto uvedieme iné tvary interpolačného polynómu, ktoré sa pohodlnejšie počítajú.

Označme

$$l_j(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_n)}{(x_j-x_0)(x_j-x_1)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_n)}, \tag{4.3}$$

pre $j = 0, 1, \dots, n$.

Zrejme $l_j \in P_n, \quad l_j(x_i) = \delta_{ij}, \quad i, j = 0, 1, \dots, n.$

Polynóm

$$p(x) = \sum_{j=0}^{n} y_j l_j(x) \quad \text{spĺňa} \quad p \in P_n \tag{4.4}$$

a naviac

$$p(x_i) = \sum_{j=0}^{n} y_j l_j(x_i) = y_i, \quad i = 0, 1, \dots, n,$$

a preto (podľa jednoznačnosti z vety 4.1) ide opäť o interpolačný polynóm. Tvar (4.4) sa nazýva *Lagrangeov tvar* interpolačného polynómu.

Príklad 4.1 Nech $x_0 = 0, x_1 = -1, x_2 = 1, y_0 = 1, y_1 = 2, y_2 = 3$. Nájdite polynóm druhého stupňa tak, aby

$$p(x_i) = y_i, \qquad i = 0, 1, 2.$$

Riešenie: Dosadením hodnôt do (4.4) máme

$$\begin{aligned} p(x) &= \frac{(x+1)(x-1)}{(0+1)(0-1)}1 + \frac{(x-0)(x-1)}{(-1-0)(-1-1)}2 + \frac{(x-0)(x+1)}{(1-0)(1+1)}3 \\ &= 1 + \frac{1}{2}x + \frac{3}{2}x^2. \end{aligned}$$

□

Označme teraz

$$\begin{aligned} p_{n-1}^{(0,n-1)} &\in P_{n-1} \quad \text{polynóm spĺňajúci} \quad p_{n-1}^{(0,n-1)}(x_i) &= y_i, \qquad i &= 0, 1, \dots, n-1, \\ p_{n-1}^{(1,n)} &\in P_{n-1} \quad \text{polynóm spĺňajúci} \quad p_{n-1}^{(1,n)}(x_i) &= y_i, \qquad i &= 1, 2, \dots, n. \end{aligned}$$

Pozrime sa teraz na polynóm

$$p(x) = \frac{1}{x_n - x_0}\left[(x_n - x)p_{n-1}^{(0,n-1)}(x) + (x - x_0)p_{n-1}^{(1,n)}(x)\right]. \tag{4.5}$$

Platí $p \in P_n$ a

$$\begin{aligned}
p(x_0) &= p_{n-1}^{(0,n-1)}(x_0) && = y_0,\\
p(x_n) &= p_{n-1}^{(1,n)}(x_n) && = y_n,\\
p(x_i) &= \frac{(x_n - x_i)p_{n-1}^{(0,n-1)}(x_i) + (x_i - x_0)p_{n-1}^{(1,n)}(x_i)}{x_n - x_0} && = y_i, \quad i = 1, 2, \ldots, n-1,
\end{aligned}$$

a preto je to opäť interpolačný polynóm. □

Označme ešte $p_k \in P_k$ polynóm spĺňajúci $p_k(x_i) = y_i = f(x_i), \quad i = 0, 1, \ldots, k$. Zrejme $p = p_n$.

Platí

$$\begin{aligned}
p_0 &= f(x_0),\\
p_1(x) &= f(x_0) + \tfrac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) = p_0(x) + c_1(x - x_0),\\
p_2(x) &= p_1(x) + c_2(x - x_0)(x - x_1),\\
&\vdots\\
p_n(x) &= p_{n-1}(x) + c_n(x - x_0)(x - x_1)\ldots(x - x_{n-1}),
\end{aligned}$$

pričom pre koeficient $c_k, \ k = 1, 2, \ldots, n$ dostávame

$$c_k = \frac{f(x_k) - p_{k-1}(x_k)}{(x_k - x_0)(x_k - x_1)\ldots(x_k - x_{k-1})}.$$

Ako vyzerá koeficient pri x^n ? Označme

$$\begin{array}{llll}
f[x_0, \ldots, x_{n-1}] & \text{koeficient pri} & x^{n-1} & \text{v polynóme} \quad p_{n-1}^{(0,n-1)},\\
f[x_1, \ldots, x_n] & \text{koeficient pri} & x^{n-1} & \text{v polynóme} \quad p_{n-1}^{(1,n)},\\
f[x_0, \ldots, x_n] & \text{koeficient pri} & x^{n} & \text{v polynóme} \quad p_{n}^{(0,n)}.
\end{array}$$

Zo vzťahu (4.5) vyplýva

$$f[x_0, \ldots, x_n] = \frac{1}{x_n - x_0}\left(f[x_1, \ldots, x_n] - f[x_0, \ldots, x_{n-1}]\right). \tag{4.6}$$

Teda

$$\begin{aligned}
f[x_0] &= f(x_0),\\
f[x_0, x_1] &= \frac{1}{x_1 - x_0}(f(x_1) - f(x_0)),\\
f[x_0, x_1, x_2] &= \frac{1}{x_2 - x_0}(f[x_1, x_2] - f[x_0, x_1]).
\end{aligned}$$

Čísla $f[x_0, x_1, \ldots, x_n]$ sa nazývajú *Newtonove pomerné diferencie.* Interpolačný polynóm v *Newtonovom tvare* je teda

$$\begin{aligned}
p(x) &= f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2]\\
&\quad + \ldots + (x - x_0)(x - x_1)\ldots(x - x_{n-1})f[x_0, \ldots, x_n].
\end{aligned} \tag{4.7}$$

Lema 4.1 *Pre pomerné diferencie platí*

$$f[x_0,\ldots,x_n]=\sum_{k=0}^{n}\frac{f(x_k)}{\prod\limits_{\substack{i=0\\ i\neq k}}^{n}(x_k-x_i)}.$$

DÔKAZ: Použijeme princíp matematickej indukcie vzhľadom na n. Pre $n=0$ platí $f[x_0]=f(x_0)$. Nech tvrdenie platí pre n. Dokážme ho pre $(n+1)$. Použijeme pritom rekurenciu (4.6).

$$\begin{aligned}
f[x_0,\ldots,x_{n+1}] &= \frac{f[x_1,\ldots,x_{n+1}]-f[x_0,\ldots,x_n]}{x_{n+1}-x_0}\\
&= \frac{1}{x_{n+1}-x_0}\left[\sum_{k=1}^{n+1}\frac{f(x_k)}{\prod\limits_{\substack{i=1\\ i\neq k}}^{n+1}(x_k-x_i)}-\sum_{k=0}^{n}\frac{f(x_k)}{\prod\limits_{\substack{i=0\\ i\neq k}}^{n}(x_k-x_i)}\right]\\
&= \frac{1}{x_{n+1}-x_0}\left(\frac{f(x_{n+1})}{\prod\limits_{i=1}^{n}(x_{n+1}-x_i)}-\frac{f(x_0)}{\prod\limits_{i=1}^{n}(x_0-x_i)}\right.\\
&\quad\left.+\sum_{k=1}^{n}\frac{f(x_k)}{\prod\limits_{\substack{i=1\\ i\neq k}}^{n}(x_k-x_i)}\left[\frac{1}{x_k-x_{n+1}}-\frac{1}{x_k-x_0}\right]\right)\\
&= \frac{f(x_0)}{\prod\limits_{i=1}^{n+1}(x_0-x_i)}+\frac{f(x_{n+1})}{\prod\limits_{i=0}^{n}(x_{n+1}-x_i)}+\sum_{k=1}^{n}\frac{f(x_k)}{\prod\limits_{\substack{i=0\\ i\neq k}}^{n+1}(x_k-x_i)}\\
&= \sum_{k=0}^{n+1}\frac{f(x_k)}{\prod\limits_{\substack{i=0\\ i\neq k}}^{n+1}(x_k-x_i)}.
\end{aligned}$$

□

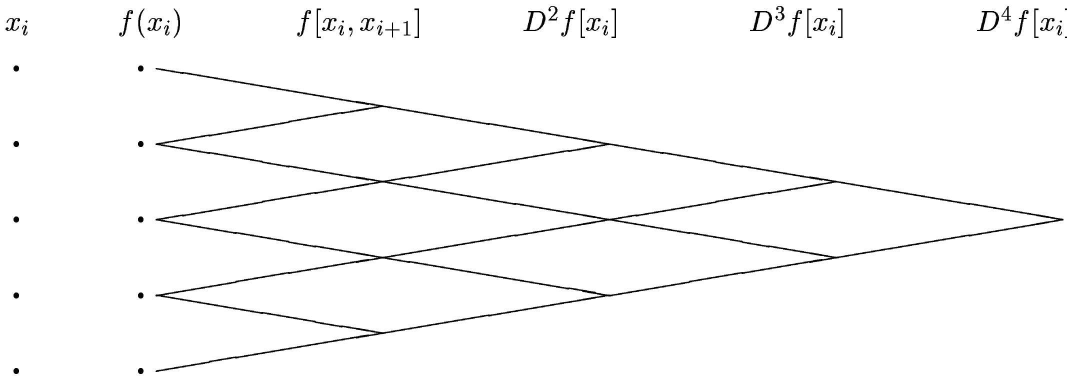
Schéma výpočtu Označme $D^r f[x_i]=f[x_i,x_{i+1},\ldots,x_{i+r}]$. Výpočet x_i, $f(x_i)$, $f[x_i,x_{i+1}]$, $D^2f[x_i]$, $D^3f[x_i]$, $D^4f[x_i],\ldots$ je schematicky znázornený na obrázku 4.1.

ALGORITMUS NA VÝPOČET $D^r f[x_0]$

Vstup $x_i, d_i=f(x_i)\quad i=0,1,2,\ldots,n$

1. Pre $i=1,2,\ldots,n$
2. Pre $j=n,n-1,\ldots,i$

   $d_j=(d_j-d_{j-1})/(x_j-x_{j-i})$
3. Výstup $d_i=f[x_0,\ldots,x_i]$



Obr. 4.1. Newtonove pomerné diferencie

ALGORITMUS NA VÝPOČET HODNOTY $p(t)$ (VARIANTA HORNEROVEJ SCHÉMY)
Pôvodná Hornerova schéma je

$$V(t) = \sum_{i=0}^{n} a_i t^i = (((a_n t + a_{n-1})t + a_{n-2})t + \ldots + a_1)t + a_0$$

Vstup $a_i, i = 0, \ldots, n; t, n$
$V = a_n$
pre $i = n-1, \ldots, 0$
$V = Vt + a_i$
výstup V

ALGORITMUS NA VÝPOČET NEWTON
Vstup $n,\ t,\ x_i,\ d_i = f[x_0, \ldots, x_i] \quad i = 0, \ldots, n$
1. $p = d_n$
2. Pre $i = n-1, n-2, \ldots, 0$
   $p = p(t - x_i) + d_i$
3. Výstup $p (= p_n(t))$ Koniec

Pozrime sa teraz na chybu, ktorej sa dopustíme, keď nahradíme f pomocou p.

Veta 4.2 *Nech $x_i, y_i \in \mathbb{R}, i = 0, 1, \ldots, n; x_i \neq x_j$ pre $i \neq j$. Označme I_t najmenší interval obsahujúci $x_0, x_1, \ldots, x_n, t$. Nech $f \in C^{n+1}(I_t)$. Potom existuje $\xi \in I_t$ také, že*

$$f(t) - p(t) = \frac{1}{(n+1)!}(t - x_0)(t - x_1)\ldots(t - x_n)f^{(n+1)}(\xi),$$

kde p je interpolačný polynóm spĺňajúci $p \in P_n,\ p(x_i) = f(x_i)$, pre $i = 0, 1, 2, \ldots, n$.

DÔKAZ:

a) Nech t je ľubovoľný, ale pevne zvolený bod taký, že $t \neq x_i$ pre $i = 0, 1, 2, \ldots, n$. Označme

$$\begin{aligned} E(x) &= f(x) - p(x), \\ \psi(x) &= (x - x_0) \ldots (x - x_n), \\ G(x) &= E(x) - \frac{\psi(x)}{\psi(t)} E(t), \quad x \in I_t. \end{aligned}$$

Funkcia G má $(n+1)$ spojitých derivácií na I_t. Platí

$$G(x_i) = 0, \quad i = 0, \ldots, n, \quad G(t) = 0,$$

t.j. G má aspoň$(n+2)$ koreňov v I_t. Preto G' má aspoň $(n+1)$ koreňov v I_t, $G^{(n+1)}$ má aspoň jeden koreň v I_t, t.j. $\exists \xi \in I_t : \ G^{(n+1)}(\xi) = 0$.

Platí

$$0 = G^{(n+1)}(\xi) = E^{(n+1)}(\xi) - \frac{\psi^{(n+1)}(\xi)}{\psi(t)} E(t) = f^{(n+1)}(\xi) - \frac{(n+1)!}{\psi(t)} E(t),$$

a teda

$$E(t) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \quad \psi(t).$$

Takže dostaneme

$$G(x) = f(x) - p(x) - \psi(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Odtiaľ vyplýva požadovaná rovnosť pre $x = t$, lebo $G(t) = 0$.

b) Nech $t = x_i$ pre nejaké i, potom teda $f(t) - p(t) = f(x_i) - p(x_i) = 0$ a platí tvrdenie vety. □

Dôsledok 4.1 *Pre všetky $t \in I_t$ platí odhad*

$$|f(t) - p(t)| \le \max_{\xi \in I_t} \left| f^{(n+1)}(\xi) \right| \left| \frac{(t - x_0)(t - x_1) \ldots (t - x_n)}{(n+1)!} \right|.$$

4.1.1 Chybová formula pomocou diferencií

Z vety 4.2 vyplýva

$$f(t) - p_n(t) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(t - x_0)(t - x_1) \ldots (t - x_n). \tag{4.8}$$

Z lemy 4.1 dostaneme

$$f[x_0, \ldots, x_n, t] = \frac{f(t)}{\prod_{i=0}^{n} (t - x_i)} + \sum_{k=0}^{n} \frac{f(x_k)}{(x_k - t) \prod_{\substack{i=0 \\ i \neq k}}^{n} (x_k - x_i)}.$$

Odtiaľ vyplýva

$$\begin{aligned} f(t) &= \sum_{k=0}^{n} f(x_k) \prod_{\substack{i=0 \\ i\neq k}}^{n} \frac{t-x_i}{x_k-x_i} + f[x_0,\ldots,x_n,t] \prod_{i=0}^{n}(t-x_i) \\ &= p_n(t) + f[x_0,\ldots,x_n,t] \prod_{i=0}^{n}(t-x_i). \end{aligned}$$

Porovnaním s (4.8) máme

$$f[x_0,x_1,\ldots,x_n,t] = \frac{1}{(n+1)!} f^{(n+1)}(\xi).$$

Preznačme $t = x_{n+1}, n+1 = m$. Dostávame

$$f[x_0,\ldots,x_m] = \frac{1}{m!} f^{(m)}(\xi), \quad \forall \xi \in I_{x_m}.$$

Poznámka 4.3 *Neplatí vo všeobecnosti*

$$\forall x \max_{x\in[a,b]} |f(x) - p_n(x)| \overset{n\to\infty}{\longrightarrow} f(x).$$

Kontrapríkladom je Rungeho funkcia $f(x) = \frac{1}{1+x^2}, -5 \le x \le 5$. *Konvergencia sa dá niekedy dosiahnuť pomocou špeciálnej voľby uzlov* $x_0, x_1, \ldots, x_n$. □

4.2 Hermiteova interpolácia

Úloha: Dané $x_i, y_i, y_i' \in \mathbb{R},\ i = 0,1,2,\ldots,n;\ x_i \neq x_j$ pre $i \neq j$. Hľadáme taký polynóm $p \in P_{2n+1}$, pre ktorý platí

$$p(x_i) = y_i,\ p'(x_i) = y_i', \qquad i = 0,1,2,\ldots,n.$$

Veta 4.3 *Nech* $x_i, y_i, y_i' \in \mathbb{R},\ i = 0,1,\ldots,n,\ x_i \neq x_j$ *pre* $i \neq j$. *Potom existuje práve jeden polynóm* $h \in P_{2n+1}$, *pre ktorý platí*

$$h(x_i) = y_i, \quad h'(x_i) = y_i'. \tag{4.9}$$

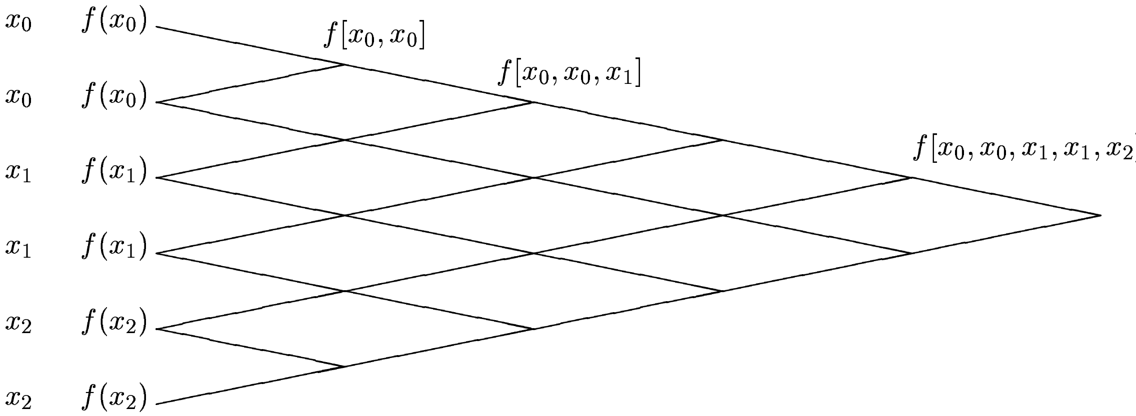
DÔKAZ:

Jednoznačnosť: Nech $h_1, h_2 \in P_{2n+1}$ spĺňajú (4.9). Potom $r = h_1 - h_2$ spĺňa $r(x_i) = r'(x_i) = 0$ pre $i = 0,1,2,\ldots,n$, a teda $x_i,\ i = 0,1,\ldots,n$ sú dvojnásobné korene. Preto $r \equiv 0$.

Existencia: Vzťahy (4.9) určujú systém $2n+2$ rovníc o $2n+2$ neznámych. Existencia vyplýva z jednoznačnosti tak ako vo vete 4.1. □

Označme

$$\begin{aligned} h_i(x) &= [1 - 2l_i'(x)(x-x_i)]\, l_i^2(x), \\ \tilde{h}_i(x) &= (x-x_i) l_i^2(x), \end{aligned}$$



Obr. 4.2. Schéma

pričom l_i sú určené vzťahom (4.3). Platí

$$
\begin{aligned}
&h_i(x_j) = \delta_{ij}, \quad h_i'(x_j) = 0, \qquad i, j = 0, \dots, n, \\
&\tilde{h}_i(x_j) = 0, \quad \tilde{h}_i'(x_j) = \delta_{ij}.
\end{aligned}
$$

Preto

$$h(x) = \sum_{i=0}^{n} y_i h_i(x) + \sum_{i=0}^{n} y_i' \tilde{h}_i(x) \tag{4.10}$$

spĺňa (4.9). Vzťah (4.10) sa nazýva Lagrangeov tvar Hermiteovho interpolačného polynómu.

4.2.1 Newtonov tvar Hermiteovho interpolačného polynómu

Podobne ako pri interpolačnom polynóme sa aj Hermiteov interpolačný polynóm dá vyjadriť pomocou pomerných diferencií v Newtonovom tvare

$$
\begin{aligned}
p_0(x) &= f(x_0), \\
p_1(x) &= f(x_0) + f'(x_0)(x - x_0), \\
p_2(x) &= f(x_0) + f'(x_0)(x - x_0) + f[x_0, x_1](x - x_0)^2, \\
p_3(x) &= f(x_0) + f'(x_0)(x - x_0) + f[x_0, x_0, x_1](x - x_0)^2 + \tfrac{f'(x_1)}{2}(x - x_0)^2(x - x_1), \\
&\vdots \\
p_n(x) &= f[x_0] + f[x_0, x_0](x - x_0) + f[x_0, x_0, x_1](x - x_0)^2 \\
&+ f[x_0, x_0, x_1, x_1](x - x_0)^2(x - x_1) \\
&+ \ldots + f[x_0, x_0, x_1, x_1, \ldots, x_{n-1}, x_{n-1}, x_n](x - x_0)^2 \ldots (x - x_{n-1})^2 \\
&+ f[x_0, x_0, \ldots, x_{n-1}, x_{n-1}, x_n, x_n](x - x_0)^2 \ldots (x - x_{n-1})^2(x - x_n).
\end{aligned}
$$

Tu sme použili označenie napr.

$$f[x_0, x_0] = f'(x_0),$$

$$f[x_0, x_0, x_1] = \tfrac{1}{x_1 - x_0}(f[x_1, x_0] - f[x_0, x_0]).$$

Príklad 4.2 Nájdite Hermiteov interpolačný polynóm tak, aby platilo

$$p^{(j)}(x_0) = f^{(j)}(x_0), \qquad 0 \le j \le k,$$

pre daný bod x_0.

Riešenie: Použijeme Taylorov rozvoj funkcie f v bode x_0 a dostaneme

$$p(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \ldots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k. \qquad \square$$

4.2.2 Chyba Hermiteovej interpolácie

Veta 4.4 *Nech sú splnené predpoklady vety 4.3 a* $f \in C^{2n+2}(I_t)$. *Potom* $\exists \xi \in I_t$

$$f(x) - h(x) = \frac{f^{(2n+2)}(\xi)}{(2n+2)!}(x - x_0)^2 \ldots (x - x_n)^2.$$

DÔKAZ:

a) Nech x je ľubovoľný, ale pevne zvolený bod taký, že $x \neq x_i$ pre $i = 0, 1, 2, \ldots, n$. Označme

$$\Omega_n(x) = (x - x_0)^2 \ldots (x - x_n)^2,$$

$$\psi(z) = f(z) - h(z) - \frac{\Omega_n(z)}{\Omega_n(x)}(f(x) - h(x)).$$

Funkcia ψ má 2-násobné korene $x_0 \ldots x_n$ a jednoduchý koreň x. To je spolu $2n+3$ koreňov. Vieme, že $\psi \in C^{(2n+2)}(I_t)$. Preto $\psi^{(2n+2)}$ má aspoň 1 koreň. Teda existuje $\xi \in I_t$ také, že $\psi^{(2n+2)}(\xi) = 0$. Preto

$$0 = \psi^{(2n+2)}(\xi) = f^{(2n+2)}(\xi) - (2n+2)!\frac{f(x) - h(x)}{\Omega_n(x)}.$$

Odtiaľ vyplýva $f(x) - h(x) = \dfrac{f^{(2n+2)}(\xi)}{(2n+2)!}\Omega_n(x)$.

b) Nech $x = x_i$ pre nejaké i. Potom teda $f(x) - h(x) = f(x_i) - h(x_i) = 0$ a tvrdenie vety platí. $\square$

4.3 Splajny

Doteraz sme pri interpolácii používali polynómy vyšších rádov. Zvoľme teraz iný prístup. Rozdeľme daný interval na väčší počet podintervalov a na každom z nich použime interpoláciu polynómami nižších rádov.

Definícia 4.1 *Nech $a = x_0 < x_1 \ldots < x_n = b$, $y_i, i = 0, \ldots, n$ sú dané čísla. Hovoríme, že funkcia $s \in C([a,b])$ je lineárny interpolačný splajn, ak $s(x_i) = y_i$, $s|_{[x_{i-1},x_i]} \in P_1$.*

Splajny sa pre svoje vlastnosti používajú pri riešení rôznych úloh. Najdôležitejšie vlastnosti si ukážeme v ďalšom texte.

Veta 4.5 (aproximácia lineárnym splajnom) *Nech $a = x_0 < x_1 \ldots < x_n = b$, $y_i \in \mathbb{R}$, $i = 0, 1, \ldots, n$, $y_i = f(x_i)$, kde $f \in C^2([a,b])$. Nech s je príslušný lineárny interpolačný splajn. Potom existuje $C > 0$ také, že*

$$\forall x \in [a,b] \quad |s(x) - f(x)| \le Ch^2,$$

kde

$$h = \max_{1 \le i \le n} |x_i - x_{i-1}|.$$

DÔKAZ: Zrejme platí pre $x \in [x_{i-1}, x_i]$

$$s(x) = f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} \left(f(x_i) - f(x_{i-1})\right).$$

Ďalej pomocou vety o strednej hodnote pre nejaké $\xi \in [x_{i-1}, x_i]$ máme

$$\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} = f'(\xi).$$

Z Taylorovho rozvoja dostaneme

$$f(x) = f(x_{i-1}) + f'(x_{i-1})(x - x_{i-1}) + f''(\theta)\frac{(x - x_{i-1})^2}{2},$$

pre nejaké $\theta \in [x_{i-1}, x]$. Teda

$$\begin{aligned} f(x) - s(x) &= (x - x_{i-1})\left[f'(x_{i-1}) - f'(\xi)\right] + f''(\theta)\frac{(x - x_{i-1})^2}{2} \\ &= (x - x_{i-1})f''(\eta)(x_{i-1} - \xi) + f''(\theta)\frac{(x - x_{i-1})^2}{2} \end{aligned}$$

pre nejaké $\eta \in [x_{i-1}, \xi]$. Odtiaľ vyplýva

$$|f(x) - s(x)| \le Ch^2.$$

□

Definícia 4.2 *Hovoríme, že reálna funkcia $s : [a,b] \to \mathbb{R}$ je kubický splajn, ak $s|_{[x_{i-1},x_i]} \in P_3$ a $s \in C^2([a,b])$. Hovoríme, že s je kubický interpolačný splajn, ak $s(x_i) = y_i$, $i = 0, \ldots, n$.*

Koeficientov je spolu $4n$. Podmienok je $\left\{\begin{array}{ll} n+1 & \text{interpolácia,} \\ n-1 & \text{spojitosť,} \\ n-1 & \text{spojitosť 1. derivácie,} \\ n-1 & \text{spojitosť 2. derivácie,} \end{array}\right.$
teda spolu $4n - 2$. Ešte treba predpísať 2 podmienky.

Konštrukcia kubického splajnu

$s \in P_3$ na $[x_i, x_{i+1}] \quad \Rightarrow \quad s'' \in P_1$ na $[x_i, x_{i+1}]$.

Označme $s''(x_i) = M_i, \; h_i = x_{i+1} - x_i$. Zrejme

$$s''(x) = \frac{1}{h_i}\left[(x_{i+1} - x)M_i + (x - x_i)M_{i+1}\right] \qquad \text{na } [x_i, x_{i+1}]$$

(s'' takto definovaná je spojitá). Integrovaním dostaneme

$$s(x) = \frac{1}{6h_i}\Big[(x_{i+1} - x)^3 M_i + (x - x_i)^3 M_{i+1}\Big] + C_i(x_{i+1} - x) + D_i(x - x_i).$$

Má platiť

$$s(x_i) = y_i, \; s(x_{i+1}) = y_{i+1} \Rightarrow C_i = \frac{y_i}{h_i} - \frac{h_i M_i}{6}, \; D_i = \frac{y_{i+1}}{h_i} - \frac{h_i M_{i+1}}{6},$$

t.j. na $[x_i, x_{i+1}]$

$$\begin{aligned} s(x) &= \frac{1}{6h_i}\left[(x_{i+1} - x)^3 M_i + (x - x_i)^3 M_{i+1}\right] \\ &+ \frac{1}{h_i}\left[(x_{i+1} - x)y_i + (x - x_i)y_{i+1}\right] \\ &- \frac{h_i}{6}\left[(x_{i+1} - x)M_i + (x - x_i)M_{i+1}\right]. \end{aligned}$$

Zatiaľ máme spojitosť, interpolačnú vlastnosť, spojitosť 2. derivácie. Ešte zostáva zabezpečiť spojitosť 1. derivácie. Na $[x_i, x_{i+1}]$ platí

$$s'(x) = \frac{1}{2h_i}\Big(-(x_{i+1} - x)^2 M_i + (x - x_i)^2 M_{i+1}\Big) + \frac{1}{h_i}(y_{i+1} - y_i) - \frac{1}{6}(M_{i+1} - M_i)h_i.$$

Na $[x_{i-1}, x_i]$ platí

$$s'(x) = \frac{1}{2h_{i-1}}\Big(-(x_i - x)^2 M_{i-1} + (x - x_{i-1})^2 M_i\Big) + \frac{y_i - y_{i-1}}{h_{i-1}} - \frac{1}{6}(M_i - M_{i-1})h_{i-1}.$$

Má platiť spojitosť. Teda po úprave

$$\frac{h_{i-1}}{6}M_{i-1} + \frac{h_i + h_{i-1}}{3}M_i + \frac{h_i}{6}M_{i+1} = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \qquad i = 1, 2, \ldots, n-1.$$

Ešte môžeme predpísať chýbajúce dve podmienky

$$s'(x_0) = y'_0, \quad s'(x_n) = y'_n,$$

t.j.

$$\begin{aligned} \frac{h_0}{3}M_0 \quad &+ \frac{h_0}{6}M_1 &&= \frac{y_1 - y_0}{h_0} - y'_0, \\ \frac{h_{n-1}}{6}M_{n-1} \quad &+ \frac{h_{n-1}}{3}M_n &&= y'_n - \frac{y_n - y_{n-1}}{h_{n-1}}. \end{aligned}$$

Máme teda $n+1$ rovníc o $n+1$ neznámych

$$\begin{pmatrix} \frac{h_0}{3} & \frac{h_0}{6} & & & & \\ \frac{h_0}{6} & \frac{h_0+h_1}{3} & \frac{h_1}{6} & & & \\ & \frac{h_1}{6} & \frac{h_1+h_2}{3} & \frac{h_2}{6} & & \\ & \vdots & & & & \\ & & & \frac{h_{n-2}}{2} & \frac{h_{n-2}+h_{n-1}}{3} & \frac{h_{n-1}}{6} \\ & & & & \frac{h_{n-1}}{6} & \frac{h_{n-1}}{3} \end{pmatrix} \begin{pmatrix} M_0 \\ \\ \\ \vdots \\ \\ M_n \end{pmatrix} = \begin{pmatrix} \frac{y_1-y_0}{h_0} & -y'_0 \\ \frac{y_2-y_1}{h_1} & -\frac{y_1-y_0}{h_0} \\ & \\ & \vdots \\ & \\ y'_n & -\frac{y_n-y_{n-1}}{h_{n-1}} \end{pmatrix}.$$

Príslušná matica je symetrická diagonálne dominantná, a teda regulárna.

Príklad 4.3 Rozhodnite, či funkcia

$$f(x) = \begin{cases} x^3 + x & \text{pre} \quad a \le x \le 0 \\ x^3 - x & \text{pre} \quad 0 < x \le b. \end{cases}$$

je kubickým splajnom.

Riešenie: Funkcia f je po častiach kubickým polynómom, pričom platí $f(0) = 0$, $\lim_{x \to 0_-} f'(x) = 1$ a $\lim_{x \to 0_+} f'(x) = -1$. Prvá derivácia teda nie je spojitá, t.j. f nie je kubickým splajnom. □

Príklad 4.4 Pre ktoré hodnoty a, b, c, d, e bude funkcia

$$f(x) = \begin{cases} a(x-2)^2 + b(x-1)^3 & \text{pre} \quad x \in (-\infty, 1] \\ c(x-2)^2 & \text{pre} \quad x \in [1, 3] \\ d(x-2)^2 + e(x-3)^3 & \text{pre} \quad x \in [3, \infty) \end{cases}$$

kubickým splajnom? Určte hodnoty a, b, c, d, e tak, aby funkcia f interpolovala body $(0, 2)$, $(3, 2)$, $(4, 6)$.

Riešenie: Výrazy $(x-1)^3$ a $(x-3)^3$ spolu aj s ich prvými i druhými deriváciami v nadpájacích bodoch sa rovnajú nule. To znamená, že na hodnotách b, e nebude záležať. Ostatné členy z definície funkcie f obsahujú vždy $(x-2)^2$. Teda ak $a = c = d$, tak f bude kubickým splajnom.

Interpolačné podmienky implikujú

$$\begin{aligned} 4a - b &= 2 \\ c &= 2 \\ 4d + e &= 6. \end{aligned}$$

Odtiaľ dostaneme $a = c = d = 2$, $b = 6$, $e = -2$. □

Vlastnosti aproximácie kubickými splajnami
Nasledujúca veta vyjadruje vlastnosť minimálnej krivosti[1].

Veta 4.6 (minimálna krivosť) *Nech*

$$\begin{aligned} g \in C^2([a,b]), \quad g'(a) &= y'_0, \quad g'(b) = y'_n, \\ g(x_i) &= y_i, \quad\quad i = 0, 1, \ldots, n. \end{aligned}$$

Nech s je príslušný kubický interpolačný splajn.
Potom

$$\int_a^b \left(s''(x)\right)^2 dx \le \int_a^b \left(g''(x)\right)^2 dx.$$

[1] Krivosť krivky $f(x)$ je daná jej druhou deriváciou $f''(x)$.

DÔKAZ: Označme $k(x) = s(x) - g(x)$. Platí

$$\int_a^b g''^2(x)dx = \int_a^b (s'' - k'')^2 dx = \int_a^b s''^2 dx - 2\int_a^b k''s''dx + \int_a^b k''^2 dx.$$

Ale

$$\begin{aligned} \int_a^b s''k''dx &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} s''k''dx = \sum_{i=0}^{n-1} \left([k's'']_{x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} k's'''dx \right) \\ &= \sum_{i=0}^{n-1} [k's'']_{x_i}^{x_{i+1}} - \sum_{i=0}^{n-1} \alpha_i \int_{x_i}^{x_{i+1}} k'dx = 0, \end{aligned}$$

lebo

$$\int_{x_i}^{x_{i+1}} k'dx = k(x_{i+1}) - k(x_i) = 0$$

a navyše

$$\begin{aligned} \sum_{i=0}^{n-1} [k's'']_{x_i}^{x_{i+1}} &= \sum_{i=0}^{n-1} [k'(x_{i+1})s''(x_{i+1}) - k'(x_i)s''(x_i)] \\ &= k'(x_n)s''(x_n) - k'(x_0)s''(x_0) \\ &= [s'(x_n) - g'(x_n)]\, s''(x_n) - [s'(x_0) - g'(x_0)]\, s''(x_0) \\ &= 0. \end{aligned}$$

□

Poznámka 4.4 *Vlastnosť minimálnej krivosti majú aj tzv. prirodzené kubické splajny, ktoré spĺňajú $s''(x_0) = s''(x_n) = 0$ a vtedy teda stačí, aby $g \in C^2([a,b])$ a $g(x_i) = y_i$ pre $i = 0, 1, 2, \ldots, n$. Potom pre prirodzený kubický splajn s tiež platí tvrdenie predchádzajúcej vety.*

O kvalite aproximácie funkcie kubickými splajnami hovorí ďalšia veta. Jej dôkaz tu neuvádzame.

Veta 4.7 (aproximačná vlastnosť) *Nech $f \in C^4([a,b])$. Nech $\mathcal{T}_n$ je postupnosť delení intervalu $[a,b]$, t.j.*

$$\mathcal{T}_n := a = x_0^{(n)} < x_1^{(n)} < \ldots < x_{n-1}^{(n)} < x_n^{(n)} = b.$$

Označme

$$|\mathcal{T}_n| = \max_{1 \le i \le n} \left| x_i^{(n)} - x_{i-1}^{(n)} \right|.$$

Nech s_n je kubický interpolačný splajn prislúchajúci funkcii f pri delení $\mathcal{T}_n$, t.j.

$$s_n \in C^2([a,b],) \qquad s_n|_{[x_i^n, x_{i+1}^n]} \in P_3,$$

$$s'(a) = f'(a), \quad s'(b) = f'(b), \quad s(x_i) = f(x_i)$$

pre $i = 0, 1, \ldots, n$. Potom existujú také konštanty C_j $(j = 0, 1, 2)$, že

$$\max_{x \in [a,b]} \left| f^{(j)}(x) - s_n^{(j)}(x) \right| \le C_j \, |\mathcal{T}_n|^{4-j} \max_{x \in [a,b]} \left| f^{(4)}(x) \right|$$

pre $j = 0, 1, 2$.

Cvičenie 4.1 Na interpoláciu funkcie e^{-x} na $[0,2]$ bol použitý polynóm 20-teho stupňa. Aký je odhad chyby interpolácie?

Cvičenie 4.2 Ukážte, že maximálna chyba lineárnej interpolácie je ohraničená číslom $\frac{M}{8}(x_2-x_1)^2$, kde $M = \max\limits_{x_1 \le x \le x_2} |f''(x)|$.

Cvičenie 4.3 Rozhodnite, či funkcia f je lineárnym splajnom, pričom

$$f(x) = \begin{cases} x & -1 \le x \le 0,5 \\ 0,5 + 2(x-0,5) & 0,5 \le x \le 2 \\ x + 1,5 & 2 \le x \le 4. \end{cases}$$

Cvičenie 4.4 Nech sú dané body $0 = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = 1$ pre $n > 1$ a hodnoty $y_i \in \mathbb{R}$, $i = 0,1,\ldots,n$. Nech $s \in C([0,1])$ je lineárny interpolačný splajn a funkcia $g \in C^1([0,1])$ spĺňa $g(x_i) = y_i$, $i = 0,1,2,\ldots,n$. Dokážte, že platí

$$\int_0^1 (s'(x))^2 dx \le \int_0^1 (g'(x))^2 dx \ .$$

Cvičenie 4.5 Zostrojte kubický splajn s uzlovými bodmi $\{0,1,2,3\}$, ktorý je kvadratický na $[0,1]$, kubický na $[1,2]$ a kvadratický na $[2,3]$.

Cvičenie 4.6 Navrhnite hodnoty parametrov a,b,c,d tak, aby s bol kubický splajn taký, že hodnota

$$\int_0^2 (s''(x))^2 dx$$

je minimálna, kde

$$s(x) = \begin{cases} 3 + x - 9x^3 & 0 \le x \le 1 \\ a + b(x-1) + c(x-1)^2 + d(x-1)^3 & 1 \le x \le 2 \end{cases} .$$

Cvičenie 4.7 Nech s je kubický splajn s uzlovými bodmi $t_1 < t_2 < \ldots < t_n$. Predpokladajme, že na intervaloch $[t_1,t_2]$, $[t_3,t_4]$ sa s redukuje na lineárny polynóm. Čo môžeme povedať o splajne s na intervale $[t_2,t_3]$?

Cvičenie 4.8 Predpokladajme, že $s(x)$ je interpolačný splajn m-tého stupňa na intervale $[a,b]$ s n uzlovými bodmi $a = t_1 < t_2 < \ldots < t_n = b$.

a) Koľko podmienok je potrebných na jednoznačné určenie splajnu s na $[a,b]$?

b) Koľko podmienok je definovaných z interpolačných podmienok v uzlových bodoch?

c) Koľko podmienok sa dá získať zo spojitosti derivácií?

d) Koľko dodatočných podmienok je ešte potrebné získať?

Kapitola 5

Metóda najmenších štvorcov

Myšlienka metódy najmenších štvorcov (MNŠ) spočíva v tom, že hľadanú funkciu (danú spojite alebo diskrétne) aproximujeme polynómom tak, aby bol "súčet štvorcov odchýliek" čo najmenší.

V ďalšom budeme uvažovať zvlášť spojitú a diskrétnu verziu MNŠ. Symbolom P_n označujeme množinu všetkých polynómov stupňa najviac n ($\leq n$). Všetky úvahy budeme robiť na intervale $[-1,1]$.

5.1 Spojitá verzia MNŠ

Majme danú funkciu $f \in C([-1,1])$. Našou úlohou je nájsť polynóm $p \in P_n$ tak, aby platilo

$$\int_{-1}^{1}(f-p)^2dx \leq \int_{-1}^{1}(f-q)^2dx \qquad \forall q \in P_n. \tag{5.1}$$

Poznámka 5.1 *Namiesto nerovnosti (5.1) by sme mohli vyžadovať nerovnosť*

$$\int_{-1}^{1}|f-p|dx \leq \int_{-1}^{1}|f-q|dx \qquad \forall q \in P_n.$$

Zdôvodnite, prečo je voľba (5.1) výhodnejšia. □

Priestor P_n je konečnorozmerný podpriestor lineárneho priestoru $C([-1,1])$. Predpisom

$$(f,g) = \int_{-1}^{1} fg\ dx \tag{5.2}$$

môžeme zaviesť v $C([-1,1])$ skalárny súčin a predpisom

$$\rho(f,g) = \left(\int_{-1}^{1}(f-g)^2dx\right)^{1/2} \tag{5.3}$$

metriku.

Nerovnosť (5.1) môžeme teraz prepísať do tvaru

$$\rho^2(f,p) \leq \rho^2(f,q) \qquad \forall q \in P_n,$$

čo je ekvivalentné s nerovnosťou

$$\rho(f,p) \le \rho(f,q) \qquad \forall q \in P_n.$$

Hľadáme teda taký prvok p z priestoru P_n, ktorý je najbližšie k danému prvku $f \in C([-1,1])$ v zmysle metriky (5.3).

Nech $\varphi_i, i = 0,1,2,\ldots,n$ je nejaká báza priestoru P_n. Potom ku každému $q \in P_n$ existujú $a_0, a_1, \ldots, a_n \in \mathbb{R}$ také, že $q = \sum_{i=0}^{n} a_i\varphi_i$. Hľadáme teda $a_0, a_1, \ldots, a_n \in \mathbb{R}$ také, aby

$$\int_{-1}^{1} \left(f - \sum_{i=0}^{n} a_i\varphi_i\right)^2 dx \le \int_{-1}^{1} \left(f - \sum_{i=0}^{n} b_i\varphi_i\right)^2 dx \qquad \forall b_0, \ldots, b_n \in \mathbb{R}.$$

Ak označíme

$$F(a_0, \ldots, a_n) = \int_{-1}^{1} \left(f - \sum_{i=0}^{n} a_i\varphi_i\right)^2 dx,$$

potom hľadáme bod minima funkcie $F : \mathbb{R}^{n+1} \to \mathbb{R}$, t.j. také $a_0, \ldots, a_n \in \mathbb{R}^{n+1}$, že

$$F(a_0, \ldots, a_n) \le F(b_0, \ldots, b_n) \qquad \forall b_0, \ldots, b_n \in \mathbb{R}^{n+1}. \tag{5.4}$$

Upravme teraz výraz, ktorým je daná funkcia F

$$\begin{aligned} F(a_0, \ldots, a_n) &= \int_{-1}^{1} \left(f - \sum_{i=0}^{n} a_i\varphi_i\right)^2 dx \\ &= \int_{-1}^{1} f^2 dx - 2\sum_{i=0}^{n} a_i \int_{-1}^{1} f\varphi_i dx + \sum_{i=0}^{n}\sum_{j=0}^{n} a_i a_j \int_{-1}^{1} \varphi_i\varphi_j dx \\ &= \mu - 2\boldsymbol{a}^T \cdot \boldsymbol{d} + \boldsymbol{a}^T \cdot \boldsymbol{C}\boldsymbol{a}, \end{aligned}$$

pričom sme označili

$\boldsymbol{C} \in \mathbb{R}^{(n+1)\times(n+1)}$	maticu s prvkami $c_{ij} = \int_{-1}^{1} \varphi_i\varphi_j dx = (\varphi_i, \varphi_j) = c_{ji}$,
$\boldsymbol{d} \in \mathbb{R}^{n+1}$	vektor s prvkami $d_i = \int_{-1}^{1} f\varphi_i dx = (f, \varphi_i)$,
$\boldsymbol{a} \in \mathbb{R}^{n+1}$	vektor s prvkami $a_0, a_1, \ldots, a_n$,
$\mu \in \mathbb{R}$	číslo $\int_{-1}^{1} f^2 dx = (f,f)$.

Zderivujme F podľa a_i

$$\frac{\partial F}{\partial a_i} = 2\left(\sum_{j=0}^{n} c_{ij}a_j - d_i\right), \quad i = 0,1,\ldots,n.$$

Nutná podmienka pre (5.4) je teda

$$\boldsymbol{C}\boldsymbol{a} = \boldsymbol{d}. \tag{5.5}$$

Lema 5.1 *Matica $\boldsymbol{C}$ je regulárna.*

DÔKAZ: Dôkaz vyplýva zo skutočnosti, že $\boldsymbol{C}$ je Gramova matica. □

Stacionárny bod funkcie F dostaneme teda riešením systému (5.5). Ukážeme teraz, že takto vypočítané $\boldsymbol{a}$ je bodom minima[1]. Na to aby stacionárny bod $\boldsymbol{a}$ bol bodom minima pre F treba vypočítať Hessovu maticu druhých parciálnych derivácií F. Tá sa zrejme rovná matici $2\boldsymbol{C}$. Táto je však pozitívne semidefinitná, lebo pre každé $x = \sum_{i=0}^{n} \alpha_i \varphi_i$ platí

$$\begin{aligned} 0 \le (x,x) &= \sum_{j=0}^{n}\sum_{i=0}^{n} \alpha_i \alpha_j (\varphi_i, \varphi_j) \\ &= (\alpha_1, \ldots, \alpha_n)\boldsymbol{C}(\alpha_1, \ldots, \alpha_n)^T. \end{aligned}$$

Pretože $\boldsymbol{C}$ je regulárna a pozitívne semidefinitná, je i pozitívne definitná. Teda $\boldsymbol{a}$ bol bodom minima pre F.

Príklad 5.1 Zvoľme bázu

$$\varphi_i = x^i,\ i = 0, \ldots, n. \tag{5.6}$$

Potom

$$\begin{aligned} c_{ij} &= \int_{-1}^{1} x^i x^j dx = \int_{-1}^{1} x^{i+j} dx \\ &= \begin{cases} 0 & i+j+1 \text{ je párne} \\ \dfrac{2}{i+j+1} & i+j+1 \text{ je nepárne} \end{cases} \end{aligned} \tag{5.7}$$

pre $i, j = 0, \ldots, n$. □

Príklad 5.2 Nech $\{\tilde{\varphi}_i\}_{i=0}^{n}$ je báza, ktorú dostaneme aplikáciou Grammovej-Schmidtovej ortonormalizácie na bázu danú pomocou (5.6). Potom $\boldsymbol{C} = \boldsymbol{I}$. V tomto prípade nemusíme systém (5.5) riešiť. Koeficienty d_i sú dané takto $d_i = \int_{-1}^{1} f\tilde{\varphi}_i dx$. Sú to vlastne Fourierove koeficienty prvku f pri báze $\tilde{\varphi}_i, i = 0, \ldots, n$

$$p_n = \sum_{i=0}^{n} (f, \tilde{\varphi}_i)\tilde{\varphi}_i.$$

Polynómy $\tilde{\varphi}_i$ sa nazývajú *Legendrove polynómy* a sú tabulované, resp. existujú podprogramy na ich výpočet. Prvých päť je

$$\tilde{\varphi}_0 = \sqrt{\frac{1}{2}}, \quad \tilde{\varphi}_1 = \sqrt{\frac{3}{2}}x, \quad \tilde{\varphi}_2 = \sqrt{\frac{5}{2}}(3x^2 - 1),$$
$$\tilde{\varphi}_3 = \sqrt{\frac{7}{2}}(5x^3 - 3x), \quad \tilde{\varphi}_4 = \sqrt{\frac{9}{2}}(35x^4 - 30x^2 + 3).$$ □

[1] Ak $\{\varphi_i\}_{i=0}^{n}$ je ortonormálna báza vzhľadom na (5.3) potom $\boldsymbol{C} = \boldsymbol{I}$ a

$$\begin{aligned} F(a_0, \ldots, a_n) &= \mu - 2\boldsymbol{a}^T \cdot \boldsymbol{d} + \boldsymbol{a}^T \cdot \boldsymbol{a} \\ &= \mu + (\boldsymbol{a} - \boldsymbol{d})^T \cdot (\boldsymbol{a} - \boldsymbol{d}) - \boldsymbol{d}^T \cdot \boldsymbol{d}\,. \end{aligned}$$

Zrejme

$$F(a_0, \ldots, a_n) \ge F(d_0, \ldots, d_n),$$

a minimum sa nadobúda pre $\boldsymbol{a} = \boldsymbol{d}$.

Poznámka 5.2 *Báza z príkladu 5.1 je nevhodná, lebo matica* $\boldsymbol{C}$ *je veľmi zle podmienená, systém (5.5) sa pri tejto báze pre* $n \geq 6$ *veľmi ťažko numericky rieši.* □

Poznámka 5.3 *Ortonormálna báza má navyše tú výhodu, že pri zvyšovaní stupňa polynómu nemusíme prepočítavať koeficienty*

$$p_{n+1} = p_n + (f, \tilde{\varphi}_{n+1})\tilde{\varphi}_{n+1}.$$ □

O konvergencii MNŠ hovorí nasledujúca veta.

Veta 5.1 *Nech* $f \in C([-1,1])$, $p_n \in P_n$ *je aproximácia funkcie* f *pomocou MNŠ, t.j. nech platí (5.1). Potom*

$$\lim_{n\to\infty} \rho(f, p_n) = 0.$$

Dôkaz: Zrejme

$$\rho(f, p_1) \geq \rho(f, p_2) \geq \ldots \geq \rho(f, p_n) \geq \ldots .$$

Nech $\varepsilon > 0$. Podľa Weierstrassovej vety sa každá spojitá funkcia na uzavretom intervale dá rovnomerne aproximovať polynómom s ľubovoľnou presnosťou, t.j.

$$\forall \varepsilon > 0\ \exists n \in \mathbb{N}\ \exists Q \in P_n \quad \forall x \in [-1,1]: \quad |f(x) - Q(x)| < \frac{\varepsilon}{\sqrt{2}}.$$

Podľa (5.1) platí

$$\rho(f, p_n) \leq \rho(f, Q) = \left[\int_{-1}^{1} (f-Q)^2 dx\right]^{1/2} \leq \left(\left(\frac{\varepsilon}{\sqrt{2}}\right)^2 .2\right)^{1/2} = \varepsilon.$$ □

5.2 Diskrétna verzia MNŠ

Majme dané $x_i, y_i \in \mathbb{R}$, $i = 0, \ldots, n$. Predpokladáme, že y_i sú hodnoty nejakej funkcie f v bodoch x_i (napríklad namerané hodnoty veličiny závislej na čase). Neznámu funkciu f chceme aproximovať polynómom $p \in P_m$ tak, aby súčet štvorcov rozdielov $f(x_i) - p(x_i)$ bol čo najmenší, t.j.

$$\sum_{i=0}^{n} (y_i - p(x_i))^2 \leq \sum_{i=0}^{n} (y_i - q(x_i))^2 \qquad \forall q \in P_m.$$

Ak by $m \geq n$, potom môžeme danými bodmi (x_i, y_i), $i = 0, \ldots, n$ preložiť hľadaný polynóm, takže sa dopustíme nulovej chyby. Preto budeme ďalej predpokladať $m < n$.

Zaveďme teraz skalárny súčin v priestore P_n predpisom

$$(f, g) = \sum_{i=0}^{n} f(x_i) g(x_i) \tag{5.8}$$

a metriku vzťahom

$$\rho(f, g) = \left(\sum_{i=0}^{n} [f(x_i) - g(x_i)]^2\right)^{1/2}. \tag{5.9}$$

Polynóm p hľadáme opäť v tvare $p = \sum_{i=0}^{m} a_i\varphi_i$, kde φ_i je báza P_m. Budeme teda minimalizovať funkciu

$$F(a_0,\dots,a_m) = \sum_{i=0}^{n}\left(y_i - \sum_{j=0}^{m} a_j\varphi_j(x_i)\right)^2.$$

Príklad 5.3 Aproximujte body (x_i, y_i) pre $i = 0,\dots,n$ lineárnou funkciou pomocou MNŠ.
Riešenie: Ak $m = 1$, bázové funkcie sú $\varphi \in \{1, x\}$. Skalárny súčin na množine dát bude

$$(\varphi_r, \varphi_s) = \sum_{i=0}^{n} \varphi_r(x_i)\varphi_s(x_i).$$

Algebraická sústava rovníc bude mať tvar

$$\begin{pmatrix} n+1 & \sum_{i=0}^{n} x_i \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \end{pmatrix}.$$

Riešenie je

$$a_1 = \frac{\sum_{i=0}^{n} x_i y_i - \dfrac{\sum_{i=0}^{n} x_i \cdot \sum_{i=0}^{n} y_i}{n+1}}{\sum_{i=0}^{n} x_i^2 - \dfrac{\left(\sum_{i=0}^{n} x_i\right)^2}{n+1}}, \qquad a_0 = \frac{\sum_{i=0}^{n} y_i}{n+1} - a_1 \frac{\sum_{i=0}^{n} x_i}{n+1}.$$

Hľadaná priamka je daná $y = a_0 + a_1 x$. □

5.3 Totálna metóda najmenších štvorcov

Pri predchádzajúcom výklade diskrétnej verzie MNŠ sme minimalizovali sumu štvorcov *vertikálnych* odchyliek $\sum_{i=0}^{n}(y_i - p(x_i))^2$ od polynómu daného stupňa. Pri totálnej metóde najmenších štvorcov predpokladáme, že x_i ani y_i nie sú presné hodnoty. Budeme sa zaoberať iba lineárnym prípadom. Našou úlohou bude nájsť taký polynóm stupňa ≤ 1 (t.j. priamku), aby súčet štvorcov vzdialeností bodov (x_i, y_i) pre $i = 0,\dots,n$ od tejto priamky bol čo najmenší.

Konštrukcia: Nech $x_i, y_i \in \mathbb{R}, i = 0,\dots,n$. Hľadáme takú priamku p v tvare

$$rx + sy - c = 0, \tag{5.10}$$

aby súčet štvorcov vzdialeností bodov (x_i, y_i) od priamky p bol minimálny.

Ak priamka (5.10) prechádza bodom $(\bar{q}_1, \bar{q}_2)$, potom $c = r\bar{q}_1 + s\bar{q}_2$. Priamka (5.10) je teda daná rovnicou

$$r(x - \bar{q}_1) + s(y - \bar{q}_2) = 0, \tag{5.11}$$

kde $(\bar{q}_1, \bar{q}_2)$ je bod ležiaci na p a (r, s) je smerový vektor normály k p.

Vzdialenosť d bodu (q_1, q_2) od priamky (5.10) je daná vzťahom

$$d^2 = \frac{(rq_1 + sq_2 - c)^2}{r^2 + s^2} = \frac{[r(q_1 - \bar{q}_1) + s(q_2 - \bar{q}_2)]^2}{r^2 + s^2}.$$

Máme teda nájsť také čísla $\bar{r}, \bar{s}, \bar{q}_1, \bar{q}_2$, aby funkcia

$$D(r, s, q_1, q_2) = \sum_{i=0}^{n} \frac{[r(x_i - q_1) + s(y_i - q_2)]^2}{r^2 + s^2}$$

nadobúdala minimálnu hodnotu, t.j.

$$D(\bar{r}, \bar{s}, \bar{q}_1, \bar{q}_2) \le D(r, s, q_1, q_2).$$

Najprv určíme $\bar{q}_1, \bar{q}_2$.

Lema 5.2 *Pre ľubovoľné $r, s \in \mathbb{R}$ také, že $r^2 + s^2 \neq 0$ a pre všetky $q_1, q_2 \in \mathbb{R}$ platí*

$$D(r, s, \bar{q}_1, \bar{q}_2) \le D(r, s, q_1, q_2),$$

kde $\bar{q}_1 = \frac{1}{n+1} \sum_{i=0}^{n} x_i, \ \bar{q}_2 = \frac{1}{n+1} \sum_{i=0}^{n} y_i.$

DÔKAZ: Označme

$$\begin{aligned} w_i &= r(x_i - q_1) + s(y_i - q_2), \ i = 0, \ldots, n, \\ z_i &= r(x_i - \bar{q}_1) + s(y_i - \bar{q}_2), \ i = 0, \ldots, n, \\ \boldsymbol{e} &= (1, \ldots, 1)^T, \ h = r(\bar{q}_1 - q_1) + s(\bar{q}_2 - q_2). \end{aligned}$$

Platí $\boldsymbol{w} = \boldsymbol{z} + h\boldsymbol{e}$,

$$\begin{aligned} (\boldsymbol{z}^T, \boldsymbol{e}) &= \sum_{i=0}^{n} z_i = \sum_{i=0}^{n} [r(x_i - \bar{q}_1) + s(y_i - \bar{q}_2)] \\ &= r\left(\sum_{i=0}^{n} x_i - \sum_{i=0}^{n} x_i\right) + s\left(\sum_{i=0}^{n} y_i - \sum_{i=0}^{n} y_i\right) = 0. \end{aligned}$$

Preto

$$\begin{aligned} D(r, s, q_1, q_2) &= \frac{1}{r^2 + s^2}\|\boldsymbol{w}\|^2 = \frac{1}{r^2 + s^2}\|\boldsymbol{z} + h\boldsymbol{e}\|^2 \\ &= \frac{1}{r^2 + s^2}\left(\|\boldsymbol{z}\|^2 + 2h(\boldsymbol{z}^T, \boldsymbol{e}) + h^2\|\boldsymbol{e}\|^2\right) \\ &= \frac{1}{r^2 + s^2}\left(\|\boldsymbol{z}\|^2 + h^2\|\boldsymbol{e}\|^2\right) \\ &= D(r, s, \bar{q}_1, \bar{q}_2) + \frac{h^2(n+1)}{r^2 + s^2} \\ &\ge D(r, s, \bar{q}_1, \bar{q}_2). \end{aligned}$$

□

Vidíme teda, že priamka p musí prechádzať bodom

$$(\bar{q}_1, \bar{q}_2) = \left(\frac{1}{n+1} \sum_{i=0}^{n} x_i, \frac{1}{n+1} \sum_{i=0}^{n} y_i\right),$$

čo je "ťažisko" bodov $(x_i, y_i), i = 0, \ldots, n$.

Teraz nájdeme normálu (r, s) priamky p.

Lema 5.3 *Nech*

$$\boldsymbol{M} = \begin{pmatrix} x_0 - \bar{q}_1 & y_0 - \bar{q}_2 \\ \vdots & \vdots \\ x_n - \bar{q}_1 & y_n - \bar{q}_2 \end{pmatrix}.$$

Nech $(\bar{r}, \bar{s})$ *je vlastný vektor zodpovedajúci najmenšiemu vlastnému číslu matice* $\boldsymbol{M}^T\boldsymbol{M}$. *Potom pre všetky* $r, s \in \mathbb{R}$, $r^2 + s^2 \neq 0$ *platí*

$$D(\bar{r}, \bar{s}, \bar{q}_1, \bar{q}_2) \leq D(r, s, \bar{q}_1, \bar{q}_2).$$

DÔKAZ: Ľahko vidieť, že

$$\begin{aligned} D(r, s, \bar{q}_1, \bar{q}_2) &= \sum_{i=0}^{n} \frac{[r(x_i - \bar{q}_1) + s(y_i - \bar{q}_2)]^2}{r^2 + s^2} = \left\| \boldsymbol{M} \frac{1}{\sqrt{r^2 + s^2}} \begin{pmatrix} r \\ s \end{pmatrix} \right\|^2 \\ &= \left(\boldsymbol{M}^T\boldsymbol{M} \frac{1}{\sqrt{r^2 + s^2}} \begin{pmatrix} r \\ s \end{pmatrix}, \frac{1}{\sqrt{r^2 + s^2}} \begin{pmatrix} r \\ s \end{pmatrix} \right). \end{aligned}$$

Euklidovská norma vektora $\frac{1}{\sqrt{r^2+s^2}} \begin{pmatrix} r \\ s \end{pmatrix}$ je rovná 1. Zaujíma nás teda

$$\min_{\|\boldsymbol{t}\|=1} \left(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{t}, \boldsymbol{t} \right).$$

Z lineárnej algebry vieme, že funkcia $g(\boldsymbol{t}) = \left(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{t}, \boldsymbol{t} \right)$ nadobúda minimum na povrchu jednotkovej gule v bode, ktorý je vlastným vektorom zodpovedajúcim najmenšiemu vlastnému číslu matice $\boldsymbol{M}^T\boldsymbol{M}$. □

Príklad 5.4 Aproximujte body (x_i, y_i), $i = 0, 1, 2$ lineárnou funkciou pomocou MNŠ i totálnej MNŠ, ak je dané

$$x = \{1, 2, 6\}, \quad y = \{2, 6, 1\}.$$

Riešenie: Uvažujme najprv totálnu MNŠ. Zrejme $(\bar{q}_1, \bar{q}_2) = (3, 3)$. Pre maticu $\boldsymbol{M}$ máme

$$\boldsymbol{M} = \begin{pmatrix} -2 & -1 \\ -1 & 3 \\ 3 & -2 \end{pmatrix}, \quad \boldsymbol{M}^T\boldsymbol{M} = \begin{pmatrix} 14 & -7 \\ -7 & 14 \end{pmatrix}.$$

Určime vlastné čísla matice $\boldsymbol{M}^T\boldsymbol{M}$, teda

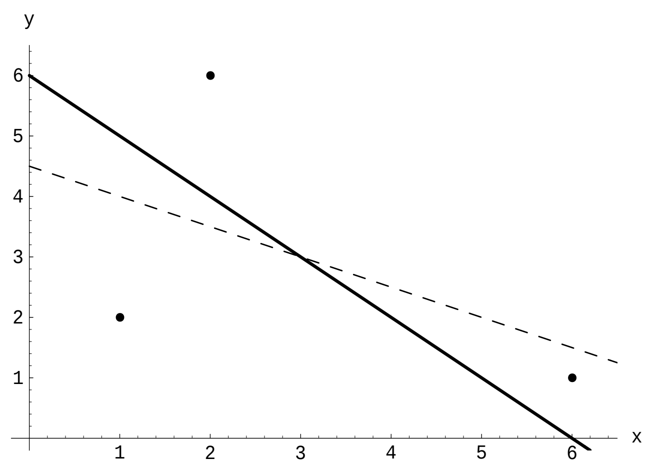
$$0 = \det(\boldsymbol{M}^T\boldsymbol{M} - \lambda\boldsymbol{I}) = \lambda^2 - 28\lambda + 147 = (\lambda - 21)(\lambda - 7).$$

Najmenšie vlastné číslo je $\lambda = 7$. Určime vlastný jednotkový vektor k nemu prislúchajúci, t.j.

$$\boldsymbol{M}^T\boldsymbol{M} \begin{pmatrix} r \\ s \end{pmatrix} = 7 \begin{pmatrix} r \\ s \end{pmatrix}, \qquad r^2 + s^2 = 1,$$

čiže

$$r = s = \frac{1}{\sqrt{2}}.$$



Obr. 5.1. Aproximácia pomocou totálnej MNŠ - plná čiara, aproximácia pomocou MNŠ - čiarkovane. Pri MNŠ je súčet druhých mocnín zvislých vzdialeností bodov od nájdenej priamky minimálny. Pri totálnej MNŠ sa minimalizuje súčet štvorcov kolmých vzdialeností bodov od priamky

Dosadením do (5.11) vidíme, že aproximácia pomocou totálnej MNŠ je priamka (pozri obrázok 5.1)

$$\frac{1}{\sqrt{2}}(x-3)+\frac{1}{\sqrt{2}}(y-3)=0 \Rightarrow y=-x+6.$$

Vzorce pre MNŠ sú uvedené v príklade 5.2. Po dosadení konkrétnych hodnôt pre x_i, y_i, $i=0,1,2$ dostaneme hľadanú priamku (viď obrázok 5.1)

$$y=-\tfrac{1}{2}x+\tfrac{9}{2}. \qquad \square$$

5.4 Metóda lineárneho krigingu

Predstavme si, že máme opísať rozmiestnenie určitej látky v istej oblasti. K dispozícii budeme mať len konečný počet meraní v nejakých náhodne zvolených miestach. Na základe výsledkov meraní by sme chceli opísať meranú veličinu v ľubovoľnom mieste oblasti. Takže na riešenie tohto problému sa prirodzene núka metóda interpolácie[2]. Doteraz sme pri interpolácii vychádzali len z jednotlivých meraní. Často však bývajú k dispozícii aj iné údaje alebo intuície o celkovom správaní sa meranej veličiny. Napr. pri výrobe zliatiny pozostávajúcej z niekoľkých zložiek vieme celkové množstvá jednotlivých komponent. Čo však nevieme, je ich priestorové rozloženie. Pevnosť zliatiny závisí od priestorového rozdelenia uhlíka C. Výrobný proces nám negarantuje úplnú homogénnosť zliatiny - tá sa dá zistiť meraniami v istých bodoch. Takže hoci priemer obsahu uhlíka v celej zliatine je konštantný, tak merania sú istým spôsobom náhodné.

Skúsme túto situáciu nejako formalizovať. Nech Ω je ohraničená oblasť v $\mathbb{R}^d$, meracie resp. pozorovacie body označme $x_1,\ldots,x_n$ a merania v nich $w_1,\ldots,w_n$. Tieto hodnoty

[2] Ak vychádzame z predpokladu, že i samotné merania majú isté chyby, tak môžeme použiť i MNŠ.

sú pre nás len jednou realizáciou náhodného poľa resp. náhodnej funkcie $w(x)$ v meracích bodoch.

Najpoužívanejšou charakteristikou náhodných veličín je ich priemerná (stredná) hodnota $E[w(x)]$. Mieru väzby dvoch náhodných veličín $w(x), w(y)$ vyjadruje ich *kovariancia*

$$\text{Cov }(w(x), w(y)) = E[(w(x) - E[w(x)])(w(y) - E[w(y)])]. \tag{5.12}$$

Rozptyl náhodnej veličiny vyjadruje vzťah

$$\text{Var }[w(x)] = E[(w(x) - E[w(x)])^2]. \tag{5.13}$$

Predpoklady
Budeme predpokladať, že náhodné pole $w(x)$ je *stacionárne*, t.j.

- $E[w(x)] = m$ pre všetky x. Bez ujmy na všeobecnosti budeme uvažovať $m = 0$.
- Miera väzby náhodných veličín v bodoch x, y bude závisieť len od ich vzdialenosti. Táto sa dá vyjadriť pomocou *kovariančnej funkcie*

$$R(|x - y|) = \text{Cov }(w(x), w(y)).$$

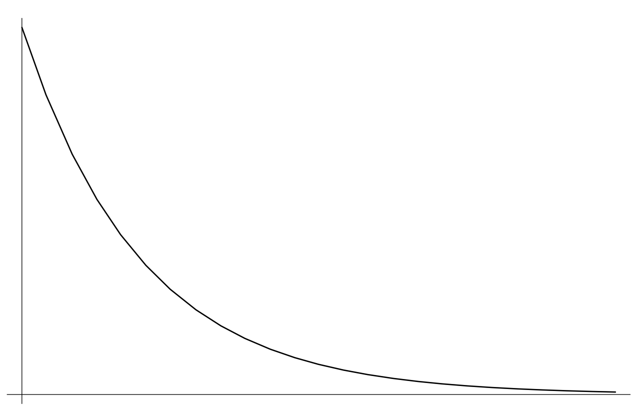
Nie každá funkcia môže byť kovariančnou funkciou. Na to musí spĺňať určité vlastnosti. Musí byť pozitívne definitná, t.j.

$$\int_0^T \psi(t) \int_0^t R(t-s)\psi(s)\; ds\; dt \geq 0 \tag{5.14}$$

pre každé $\psi \in C([0,\infty))$ a každé $T > 0$. Dá sa dokázať, že postačujúcou podmienkou pre pozitívnu definitnosť je (pozri [15])

$$(-1)^j R^{(j)}(t) \geq 0 \qquad \forall t \geq 0, \qquad j = 0, 1, 2;\; R' \neq 0. \tag{5.15}$$

Stačí teda ak funkcia bude kladná, klesajúca a konvexná - pozri obrázok 5.2. Klesajúcosť funkcie znamená, že čím ďalej sú od seba dva body, tým menej sa ovplyvňujú.



Obr. 5.2. Kovariančná funkcia

Chceli by sme vyjadriť približnú hodnotu náhodného poľa w v ľubovoľnom bode x^* pomocou $w_1, \ldots, w_n$. Budeme teda hľadať w^* v tvare

$$w^* = \sum_{j=1}^{n} \lambda_j w_j. \tag{5.16}$$

Hodnoty λ_j pre $j = 1, \ldots, n$ musia nejako závisieť od náhodného poľa w. Chceme zachovať priemer $E[w^*] = E[w(x)] = m$ bez ohľadu na hodnotu m, teda musí platiť

$$\sum_{j=1}^{n} \lambda_j = 1.$$

Ďalej chceme minimalizovať Var $[w^* - w(x^*)]$. O riešení tohto problému hovorí nasledujúca veta.

Veta 5.2 (metóda lineárneho krigingu) *Nech $E[w(x)] = 0$ pre všetky x a nech w^* je tvaru (5.16). Potom riešenie systému*

$$\left\{ \begin{array}{ll} \sum_{j=1}^{n} R(|x_k - x_j|)\lambda_j + \mu & = R(|x_k - x^*|) \quad \forall k \\ \sum_{j=1}^{n} \lambda_j & = 1 \end{array} \right. \tag{5.17}$$

s váhami λ_j a Lagrangeovým multiplikátorom μ, má minimálny rozptyl

$$Var\ [w^* - w(x^*)].$$

DÔKAZ: Jednoduchým výpočtom dostaneme

$$\frac{1}{2}\mathrm{Var}\ [w^* - w(x^*)] = \frac{1}{2}E\left[\sum_{j}^{n} \lambda_j w_j - w(x^*)\right]^2$$

$$= \frac{1}{2}\sum_{k,j=1}^{n} \lambda_k \lambda_j \mathrm{Cov}\ (w_k, w_j) - \sum_{j=1}^{n} \lambda_j \mathrm{Cov}\ (w_j, w(x^*)) + \frac{1}{2}\mathrm{Var}\ [w(x^*)].$$

Pretože

$$\mathrm{Cov}\ (w_k, w_j) = R(|x_k - x_j|),$$

máme minimalizovať výraz

$$\frac{1}{2}\sum_{k,j=1}^{n} \lambda_k \lambda_j R(|x_k - x_j|) - \sum_{j=1}^{n} \lambda_j R(|x_j - x^*|).$$

Nech $\boldsymbol{A}$ je kovariančná matica

$$\boldsymbol{A} = \begin{pmatrix} R(0) & R(|x_1 - x_2|) & R(|x_1 - x_3|) & \ldots & R(|x_1 - x_n|) \\ R(|x_2 - x_1|) & R(0) & R(|x_2 - x_3|) & \ldots & R(|x_2 - x_n|) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdot\cdot & \cdots\cdots \\ R(|x_n - x_1|) & R(|x_n - x_2|) & R(|x_n - x_3|) & \ldots & R(0) \end{pmatrix}.$$

Dá sa dokázat', že $\boldsymbol{A}$ je symetrická a pozitívne definitná (pozri [10] veta 7.1). Označme vektor $\boldsymbol{\lambda}$ s komponentami λ_j a vektor $\boldsymbol{b}$ so zložkami $R(|x_j - x^*|)$ pre $j = 1, \dots, n$. Teda máme minimalizovat' kvadratický funkcionál

$$F(\boldsymbol{\lambda}) = \frac{1}{2}(\boldsymbol{\lambda}^T, \boldsymbol{A}\boldsymbol{\lambda}) - (\boldsymbol{b}^T, \boldsymbol{\lambda}).$$

Jeho minimum je riešením systému (pozri vetu 7.5)

$$\boldsymbol{A}\boldsymbol{\lambda} = \boldsymbol{b}. \tag{5.18}$$

Nech $\boldsymbol{B}$ je matica systému (5.17)

$$\boldsymbol{B} = \begin{pmatrix} R(0) & R(|x_1 - x_2|) & R(|x_1 - x_3|) & \dots & R(|x_1 - x_n|) & 1 \\ R(|x_2 - x_1|) & R(0) & R(|x_2 - x_3|) & \dots & R(|x_2 - x_n|) & 1 \\ \dots\dots\dots\dots & \dots\dots\dots\dots & \dots\dots\dots\dots & .. & \dots\dots\dots\dots & . \\ R(|x_n - x_1|) & R(|x_n - x_2|) & R(|x_n - x_3|) & \dots & R(0) & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{pmatrix}.$$

L'ahko sa overí, že $\boldsymbol{B}$ je pozitívne semidefinitná na množine všetkých vektorov dĺžky $n+1$ s nulovou poslednou zložkou, t.j.

$$(x_1, \dots, x_n, 0)\boldsymbol{B}(x_1, \dots, x_n, 0)^T = (x_1, \dots, x_n)\boldsymbol{A}(x_1, \dots, x_n)^T \geq 0.$$

Pretože $\boldsymbol{B}$ je regulárna, je i pozitívne definitná na množine všetkých vektorov dĺžky $n+1$ s nulovou poslednou zložkou. Takže jediné riešenie (5.17) má $\mu = 0$. Toto riešenie zároveň spĺňa (5.18), a teda $\boldsymbol{\lambda}$ má minimálny rozptyl Var $[w^* - w(x^*)]$. □

Veta 5.3 (ekvivalentná formulácia) *Nech w^* je tvaru (5.16) a λ_j sú dané (5.17). Potom*

$$w^* = \sum_{j=1}^{n} R(|x^* - x_j|)\alpha_j + \beta, \tag{5.19}$$

pričom α_j $\forall j, \beta$ sú riešením systému

$$\begin{cases} \displaystyle\sum_{j=1}^{n} R(|x_k - x_j|)\alpha_j + \beta & = w_k \quad \forall k \\ \displaystyle\sum_{j=1}^{n} \alpha_j & = 0. \end{cases} \tag{5.20}$$

DÔKAZ: Všimnime si, že (5.17) a (5.20) majú tú istú maticu $\boldsymbol{B}$. Tá je symetrická a regulárna. Označme jej inverznú maticu

$$\boldsymbol{B}^{-1} = (b_{i,j})_{i,j=1,\dots,n+1}.$$

Platí

$$\begin{cases} \lambda_j & = b_{j,n+1} + \displaystyle\sum_{i=1}^{n} b_{j,i} R(|x_i - x^*|), \\ \mu & = b_{n+1,n+1} + \displaystyle\sum_{i=1}^{n} b_{n+1,i} R(|x_i - x^*|), \end{cases}$$

a

$$\begin{cases} \alpha_j &= \sum_{i=1}^{n} b_{j,i} w_i, \\ \beta &= \sum_{i=1}^{n} b_{n+1,i} w_i. \end{cases}$$

Odtiaľ vyplýva

$$\begin{aligned} w^* &= \sum_{j=1}^{n} \lambda_j w_j = \sum_{j=1}^{n} w_j \left[b_{j,n+1} + \sum_{i=1}^{n} b_{j,i} R(|x_i - x^*|) \right] \\ &= \sum_{i=1}^{n} R(|x_i - x^*|) \sum_{j=1}^{n} b_{j,i} w_j + \sum_{j=1}^{n} b_{j,n+1} w_j \\ &= \sum_{i=1}^{n} R(|x_i - x^*|) \alpha_i + \beta. \end{aligned}$$

□

Z vety 5.3 vidieť, že metóda lineárneho krigingu je vlastne interpolačná metóda. Namerané hodnoty zostávajú zachované. Interpolovaná hodnota v ľubovoľnom bode x^* sa určí vzťahom (5.19). Podrobnejší popis metódy i s aplikáciami sa dá nájsť v [9]. Metóda lineárneho krigingu sa dá modifikovať i pre prípad, že namerané hodnoty sú dané s istými chybami merania. Týmto sa tu však nebudeme zaoberať.

Cvičenie 5.1 Dokážte, že predpisom (5.2), (5.3) je skutočne daný skalárny súčin a metrika v $C([-1,1])$.

Cvičenie 5.2 Preneste celú teóriu spojitej verzie MNŠ na interval $[a,b]$. Ako budú vyzerať Legendrove polynómy na $[a,b]$?

Cvičenie 5.3 Overte, že vzorcami (5.8), (5.9) sú definované skalárny súčin a metrika v P_n.

Cvičenie 5.4 Zopakujte postup zo spojitej verzie MNŠ pre diskrétny prípad.

Cvičenie 5.5 Aproximujte body $(2,3),(3,7),(7,2)$ lineárnou funkciou pomocou **totálnej metódy najmenších štvorcov**.

Kapitola 6

Riešenie $f(x) = 0$. Hl'adanie koreňov nelineárnych rovníc

Budeme sa zaoberať úlohou:

K danej spojitej funkcii $f : M \to \mathbb{R}, M \subset \mathbb{R}$ *nájsť bod* $\bar{x} \in M$ *tak, že* $f(\bar{x}) = 0$.

V závislosti na f môže naša úloha mať viac (prípadne nekonečne veľa) riešení, jedno riešenie, prípadne žiadne riešenie. My budeme predpokladať, že existuje aspoň jedno riešenie a budeme sa snažiť "nájsť ho".

Je známe, že aj keď má f pomerne jednoduchý tvar (napr. polynóm) nie je možné korene analyticky vyjadriť (napr. vyjadriť korene ako funkciu koeficientov polynómu). Preto sme odkázaní na použitie približných metód. To znamená, že sa budeme snažiť skonštruovať postupnosť $\{x_n\}_{n=0}^{\infty}$, pre ktorú platí $x_n \to \bar{x}$, pričom $f(\bar{x}) = 0$.

Základným nástrojom na konštrukciu približných metód na riešenie $f(x) = 0$ sú vety o pevnom bode. Úlohu $f(x) = 0$ nahradíme úlohou hľadania pevného bodu funkcie g, t.j. úlohou:

K danej funkcii g *nájsť* $\bar{x}$ *tak, aby* $\bar{x} = g(\bar{x})$.

Funkciu g pritom volíme tak, aby

$$\bar{x} = g(\bar{x}) \Rightarrow f(\bar{x}) = 0 \tag{6.1}$$

Príklad 6.1 Ak zvolíme $g(x) = x - f(x)$, potom platí (6.1) .

6.1 Veta o pevnom bode

Veta 6.1 (Brouwerova) *Nech* $a, b \in \mathbb{R}, a < b$, $g \in C([a, b])$, $g([a, b]) \subset [a, b]$. *Potom existuje* $\bar{x} \in [a, b]$ *tak, že* $\bar{x} = g(\bar{x})$.

DÔKAZ: Označme $\mathcal{G}(x) = x - g(x)$. Platí $\mathcal{G}(a) = a - g(a) \leq 0$, $\mathcal{G}(b) = b - g(b) \geq 0$. Preto existuje $\bar{x} \in [a, b]: \quad \mathcal{G}(\bar{x}) = 0$, a teda $\bar{x} - g(\bar{x}) = 0$. □

Veta 6.2 (Banachova) *Nech* $a, b \in \mathbb{R}, a < b$ *a nech* $g : [a, b] \to [a, b]$ *spĺňa*

$$\exists 0 < \lambda < 1 \quad \forall x, y \in [a, b] : \ |g(x) - g(y)| \leq \lambda |x - y|. \tag{6.2}$$

Potom

i) existuje práve jedno $\bar{x} \in [a, b]$: $\bar{x} = g(\bar{x})$,

ii) postupnosť $x_{n+1} = g(x_n)$ $n = 0, 1, 2, \ldots$ *splňa* $x_n \to \bar{x}$ *pre ľubovoľne zvolené* $x_0 \in [a, b]$,

iii) platí odhad

$$|x_n - \bar{x}| \leq \frac{\lambda^n}{1-\lambda}|x_1 - x_0|. \tag{6.3}$$

DÔKAZ: Z (6.2) vyplýva, že g je spojitá, a teda podľa vety 6.1 existuje $\bar{x} \in [a, b]$ také, že $\bar{x} = g(\bar{x})$. Jednoznačnosť vyplýva z (6.2), pretože $0 < \lambda < 1$. Počítajme

$$|\bar{x} - x_n| = |g(\bar{x}) - g(x_{n-1})| \leq \lambda|\bar{x} - x_{n-1}| \leq \lambda^2|\bar{x} - x_{n-2}| \leq \ldots \leq \lambda^n|\bar{x} - x_0|. \tag{6.4}$$

Keďže $\lambda < 1$, platí $|\bar{x} - x_n| \to 0$, a teda platí *i), ii).*
Ďalej

$$|\bar{x} - x_0| \leq |\bar{x} - x_1| + |x_1 - x_0| = |g(\bar{x}) - g(x_0)| + |x_1 - x_0| \leq \lambda|\bar{x} - x_0| + |x_1 - x_0|,$$

a teda (keďže $0 < \lambda < 1$)

$$|\bar{x} - x_0| \leq \frac{1}{1-\lambda}|x_1 - x_0|. \tag{6.5}$$

Z (6.5) a (6.4) dostávame (6.3). □

Poznámka 6.1 *Chybu n-tej aproximácie môžeme odhadnúť aj takto*

$$|\bar{x} - x_n| \leq \frac{\lambda}{1-\lambda}|x_n - x_{n-1}|.$$

DÔKAZ: *Zrejme*

$$|\bar{x} - x_{n-1}| \leq |\bar{x} - x_n| + |x_n - x_{n-1}| \leq \lambda|\bar{x} - x_{n-1}| + |x_n - x_{n-1}|,$$

t.j.

$$|\bar{x} - x_{n-1}| \leq \frac{1}{1-\lambda}|x_n - x_{n-1}|,$$

a teda

$$|\bar{x} - x_n| = |g(\bar{x}) - g(x_{n-1})| \leq \lambda|\bar{x} - x_{n-1}| \leq \frac{\lambda}{1-\lambda}|x_n - x_{n-1}|.$$ □

Metóda postupných aproximácií $x_{n+1} = g(x_n), n = 0, 1, \ldots$ spomínaná v Banachovej vete o pevnom bode je základnou jednobodovou iteračnou metódou. Tvar funkcie g pritom rozhoduje o konvergencii metódy - pozri obrázky 6.1 a 6.2.

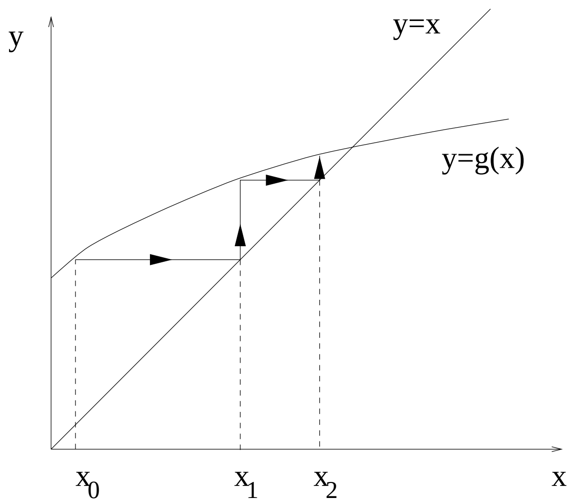
Konštantu λ, ktorá vystupuje v (6.2) môžeme odhadnúť pomocou vety o strednej hodnote (ak $g \in C^1([a, b])$). O tom hovorí nasledujúca veta.

Veta 6.3 *Nech* $g \in C^1([a, b]), g([a, b]) \subset [a, b]$ *a*

$$\lambda := \max_{[a,b]} |g'(x)| < 1.$$

Potom platia všetky tvrdenia vety 6.2 a naviac platí

$$\lim_{n\to\infty} \frac{\bar{x} - x_{n+1}}{\bar{x} - x_n} = g'(\bar{x}). \tag{6.6}$$



Obr. 6.1. Metóda postupných aproximácií konverguje

DÔKAZ: Z vety o strednej hodnote vyplýva

$$|g(x) - g(y)| \leq |g'(\xi)||x - y| \leq \lambda |x - y|,$$

čiže platí veta 6.2. Ďalej platí

$$\bar{x} - x_{n+1} = g(\bar{x}) - g(x_n) = g'(\xi_n)(\bar{x} - x_n),$$

a teda

$$\frac{\bar{x} - x_{n+1}}{\bar{x} - x_n} = g'(\xi_n), \tag{6.7}$$

kde ξ_n leží medzi $\bar{x}, x_n$. Zrejme $\xi_n \to \bar{x}$ a zo spojitosti g' vyplýva $g'(\xi_n) \to g'(\bar{x})$. Platí teda (6.6). □

Poznámka 6.2 *Nech $|g'(\bar{x})| > 1$. Potom postupnosť $x_{n+1} = g(x_n)$ nekonverguje k $\bar{x}$, lebo*

$$|x_{n+1} - \bar{x}| = |g(x_n) - g(\bar{x})| = |g'(\xi_n)||x_n - \bar{x}| > |x_n - \bar{x}|$$

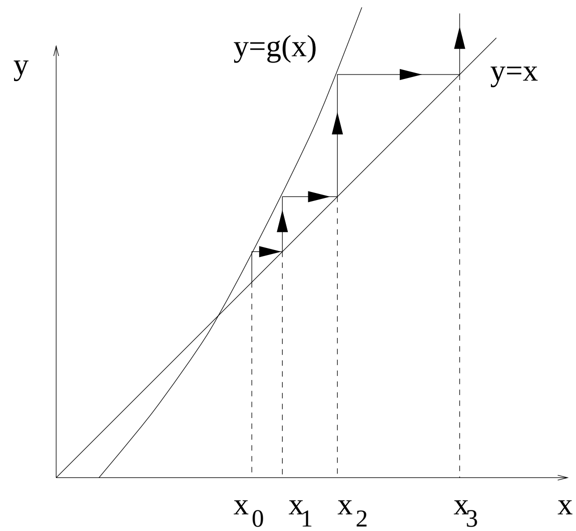
(predpokladali sme, že g' je spojitá, preto $g'(\xi_n) > 1$ pre veľké hodnoty n). □

V nasledujúcej vete sa zbavíme predpokladu $g([a, b]) \subset [a, b]$, ktorý nie je vždy ľahké overiť.

Veta 6.4 *Nech $\bar{x} = g(\bar{x})$, nech g je spojite diferencovateľná v okolí $O(\bar{x})$ bodu $\bar{x}$, $|g'(\bar{x})| < 1$. Potom ak x_0 je dostatočne blízko $\bar{x}$ tak platí veta 6.3.*

DÔKAZ: Keďže $g \in C^1(O(\bar{x}))$ existuje taký uzavretý interval $I := [\bar{x} - \varepsilon, \bar{x} + \varepsilon]$, že $\lambda = \max_{x \in I} |g'(x)| < 1$. Ukážeme, že $g(I) \subset I$. Počítajme pre $x \in I$

$$|\bar{x} - g(x)| = |g(\bar{x}) - g(x)| \leq \lambda |\bar{x} - x| < |\bar{x} - x| \leq \varepsilon.$$



Obr. 6.2. Metóda postupných aproximácií diverguje

Preto $x \in I \Rightarrow g(x) \in I$, t.j. $g(I) \subset I$. Na I sú teda splnené predpoklady vety 6.3. $\square$

Príklad 6.2 Vypočítajte

$$\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{\cdots}}}}\,.$$

Riešenie: Daný výraz sa dá formálne vypočítať pomocou rekurzie

$$x_{k+1} = \sqrt{2+x_k}, \qquad x_0 = 0, \qquad k \geq 0.$$

Zobrazenie $f(x) = \sqrt{2+x}$ zobrazuje interval $[0,2]$ do seba, lebo $0 \leq x \leq 2 \Rightarrow \sqrt{2+x} \leq 2$. Funkcia f je spojite diferencovateľná a platí

$$\max_{x\in[0,2]} |f'(x)| = \max_{x\in[0,2]} \frac{1}{2}\frac{1}{\sqrt{2+x}} < 1.$$

Banachova veta o pevnom bode hovorí, že existuje jediný pevný bod spĺňajúci

$$x = \sqrt{2+x}, \text{ alebo } x^2 - x - 2 = 0.$$

Pozitívne riešenie je $x = 2$. $\square$

Príklad 6.3 Navrhnite iteračnú metódu na určenie nulových bodov funkcie

$$f(x) = x^2 - \ln x - 2, \qquad x > 1.$$

Použite metódu postupných aproximácií a určte vhodný interval $[a,b] \subset (0,\infty)$ tak, aby iteračná postupnosť v ňom konvergovala ku nulovému bodu funkcie f.

Riešenie: Prepíšme rovnicu $f(x) = x^2 - \ln x - 2 = 0$ do tvaru $x = \sqrt{\ln x + 2}$. Funkcia $g(x) = \sqrt{\ln x + 2}$ je spojite diferencovateľná, zobrazuje interval $[1,2]$ do seba a je v $[1,2]$ rastúca. Ďalej

$$\max_{x\in[1,2]} |g'(x)| = \max_{x\in[1,2]} \frac{1}{2x}\frac{1}{\sqrt{\ln x + 2}} < 1.$$

Použitím Banachovej vety o pevnom bode vidíme, že postupnosť $\{x_k\}_{k\geq 0}$, daná iteračným predpisom $x_{k+1} = g(x_k)$, $x_0 \in [1,2]$, konverguje k jedinému riešeniu rovnice $f(x) = x$. Jednoznačnosť vyplýva z monotónnosti funkcie g. $\square$

Príklad 6.4 Uvažujme rovnicu $x^2 = 3$, t.j. $f(x) = x^2 - 3 = 0$. Túto rovnicu prepíšeme na tvar $x = g(x)$ tromi rôznymi spôsobmi:

$$\begin{array}{lll} \text{A:} & x = x + x^2 - 3, & \text{t.j. } g(x) = x + x^2 - 3, \\ \text{B:} & x = \frac{3}{x}, & \text{t.j. } g(x) = \frac{3}{x}, \\ \text{C:} & x = \frac{1}{2}(x + \frac{3}{x}), & \text{t.j. } g(x) = \frac{1}{2}(x + \frac{3}{x}). \end{array} \tag{6.8}$$

Vo všetkých 3 prípadoch platí (6.1).

Zvoľme $x_0 = 2$ a počítajme $x_{n+1} = g(x_n)$ pre každý z prípadov A, B, C. Dostávame

	A	B	C
x_0	2	2	2
x_1	3	1,5	1,75
x_2	9	2	1,732143
x_3	87	1,5	1,732051$\approx \sqrt{3}$

.

Vidíme, že vlastnosť $x_n \to \bar{x}$ dostávame iba v prípade C. Pozrime sa teraz na $|g'(\bar{x})|$

$$\begin{array}{llll} \text{v prípade A :} & |g'(\bar{x})| & = |1 + 2\bar{x}| & = 1 + 2\sqrt{3} > 1, \\ \text{v prípade B :} & |g'(\bar{x})| & = |\frac{-3}{\bar{x}^2}| & = 1, \\ \text{v prípade C :} & |g'(\bar{x})| & = \frac{1}{2}|1 - \frac{3}{\bar{x}^2}| & = 0. \end{array}$$

$\square$

Teraz si ukážeme ako k danej funkcii f zostrojiť funkciu g tak, aby platilo (6.1) a navyše boli splnené predpoklady vety 6.2.

Veta 6.5 *Nech $f \in C^1([a,b]), f(a)f(b) < 0$ a nech existujú $K > 0, K_1 > 0$ také, že pre všetky $x \in [a,b]$ platí $K < f'(x) < K_1$. Zvoľme*

$$\varepsilon \in \left(0, \frac{K}{K + K_1}\right), \qquad M \in \left(\frac{-1+\varepsilon}{K_1}, \frac{-\varepsilon}{K}\right).$$

Potom $g(x) = x + Mf(x)$ spĺňa predpoklady vety 6.2.

DÔKAZ: Najprv ukážeme, že interval, v ktorom volíme M, je dobre definovaný, t.j. $\dfrac{-1+\varepsilon}{K_1} < \dfrac{-\varepsilon}{K}$. Keďže $0 < \varepsilon < \dfrac{K}{K + K_1}$, platí $0 < \varepsilon K + \varepsilon K_1 < K$, a teda $\varepsilon K_1 < K(1-\varepsilon)$. Odtiaľ vyplýva $\dfrac{\varepsilon}{K} < \dfrac{1-\varepsilon}{K_1}$, a teda $\dfrac{-1+\varepsilon}{K_1} < \dfrac{-\varepsilon}{K}$.
Funkcia f je striktne rastúca, čiže existuje jediné $\bar{x} \in [a,b] : f(\bar{x}) = 0$.

Pre ľubovoľné $\xi \in [a,b]$ platí $0 < K \leq f'(\xi) \leq K_1$ odkiaľ (keďže $M < 0$) vyplýva

$$\begin{array}{lllll} \varepsilon - 1 & < MK_1 & \leq Mf'(\xi) & \leq MK & < -\varepsilon, \\ \varepsilon & < 1 + MK_1 & \leq 1 + Mf'(\xi) & \leq 1 + MK & < 1 - \varepsilon < 1, \end{array}$$

čiže $\varepsilon < g'(\xi) = 1 + Mf'(\xi) < 1$, teda funkcia g je striktne rastúca na $[a,b]$.

Počítajme

$$\begin{aligned} g(x)-\bar{x} &= g(x)-g(\bar{x}) && = x+Mf(x)-(\bar{x}+Mf(\bar{x})) \\ &= x-\bar{x}+M(f(x)-f(\bar{x})) && = (x-\bar{x})(1+Mf'(\xi_1)), \end{aligned}$$

kde ξ_1 leží medzi $x, \bar{x}$.

Ukážme $g([a,b]) \subset [a,b]$. Presnejšie povedané pre $x \in [a,\bar{x}]$ ukážeme $g(x) \in [a,\bar{x}]$. Podobne sa dôkaz robí pre $x \in [\bar{x},b]$.

$$g(x)-\bar{x} = (x-\bar{x})(1+Mf'(\xi_1)) \le 0,$$

t.j.

$$g(x) \le \bar{x}.$$

Z druhej strany

$$\bar{x}-g(x) = (\bar{x}-x)(1+Mf'(\xi_1)) \le \bar{x}-x,$$

teda

$$x \le g(x)$$

platí pre všetky $x \in [a,\bar{x}]$, tým skôr pre $x=a$. Týmto sme ukázali

$$a \le g(a) \le g(x) \le \bar{x}$$

pre $x \in [a,\bar{x}]$. Navyše

$$\begin{aligned} |g(x)-g(y)| &= |x+Mf(x)-(y+Mf(y))| && = |x-y+M(f(x)-f(y))| \\ &= |x-y||1+Mf'(\xi)| && < |1-\varepsilon||x-y|, \end{aligned}$$

t.j. platí (6.2). $\square$

6.2 Newtonova metóda

Teraz si ukážeme rafinovanejšiu voľbu funkcie g. Označme

$$g(x) = x - \frac{f(x)}{f'(x)}. \tag{6.9}$$

Postupnosť x_n bude mať teda tvar

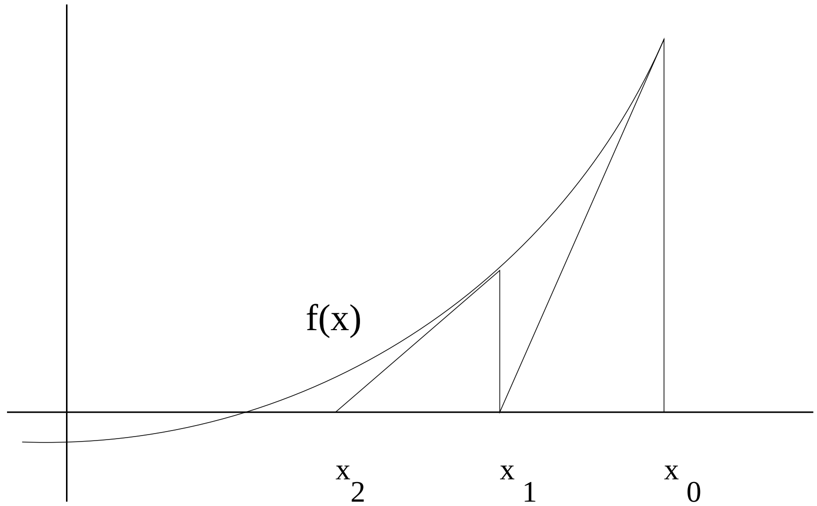
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{6.10}$$

Táto metóda sa nazýva *Newtonova metóda.* Graficky to znamená, že x_{n+1} volíme ako koreň dotyčnice grafu funkcie f, ktorá prechádza bodom $[x_n, f(x_n)]$.

O konvergencii Newtonovej metódy hovorí nasledujúca veta.

Veta 6.6 *Nech $\bar{x} \in \mathbb{R}, f \in C^2(O(\bar{x})), f(\bar{x}) = 0, f'(x) \neq 0$. Potom existuje také okolie $O_1(\bar{x})$, že postupnosť (6.10) spĺňa $x_n \to \bar{x}$. Naviac platí*

$$\lim_{n\to\infty} \frac{x_{n+1}-\bar{x}}{(x_n-\bar{x})^2} = \frac{g''(\bar{x})}{2}. \tag{6.11}$$



Obr. 6.3. Newtonova metóda

DÔKAZ: Pre funkciu g definovanú pomocou (6.9) platí

$$g'(\bar{x}) = 1 - \frac{f'(\bar{x})f'(\bar{x}) - f(\bar{x})f''(\bar{x})}{f'(\bar{x})^2} = 0.$$

Teda sú splnené predpoklady vety 6.4. Ďalej z Taylorovej vety vyplýva

$$\begin{aligned} x_{n+1} &= g(x_n) = g(\bar{x}) + g'(\bar{x})(x_n - \bar{x}) + g''(\xi_n)\frac{(x_n - \bar{x})^2}{2} \\ &= \bar{x} + \frac{1}{2}g''(\xi_n)(x_n - \bar{x})^2. \end{aligned}$$

Teda pre nejaké ξ_n medzi $x_n, \bar{x}$ platí

$$\frac{x_{n+1} - \bar{x}}{(x_n - \bar{x})^2} = \frac{1}{2}g''(\xi_n). \tag{6.12}$$

Limitným prechodom dostávame (6.11). □

Jednoduchý príklad kedy Newtonova iteračná metóda nemusí konvergovať je znázornený na obrázku 6.4.

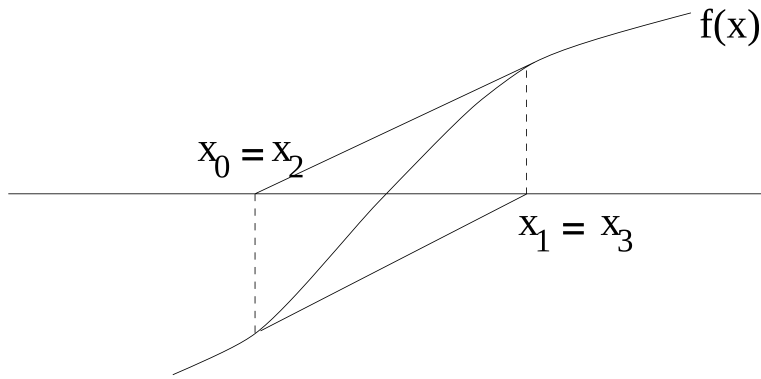
Vidíme teda, že v prípade jednoduchého koreňa (predpoklad $f'(\bar{x}) \neq 0$) konverguje Newtonova metóda rýchlejšie ako všeobecná metóda tvaru $x_{n+1} = g(x_n)$. Toto dostaneme porovnaním (6.7) a (6.12), lebo pre Newtonovu metódu dostaneme

$$|x_{n+1} - \bar{x}| \leq C|x_n - \bar{x}|^2,$$

kým pre všeobecnú metódu platí

$$|x_{n+1} - \bar{x}| \leq C|x_n - \bar{x}|.$$

Nasledujúca veta hovorí ako treba modifikovať Newtonovu metódu pre prípad p-násobného koreňa.



Obr. 6.4. Zacyklené iterácie

Veta 6.7 *Nech $p \in \mathbb{N}$, $\bar{x}$ je p-násobným koreňom funkcie $f \in C^2(O(\bar{x}))$ a $f'(x)$ je nenulová v $O(\bar{x}) \setminus \{\bar{x}\}$. Nech $f^{(p)}(\bar{x}) \neq 0$.*
Potom postupnosť

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

spĺňa

$$x_n \to \bar{x}, \qquad \lim_{n\to\infty} \frac{x_{n+1}-\bar{x}}{(x_n-\bar{x})^2} = C.$$

DÔKAZ: Budeme postupovať podobne ako v dôkaze predošlej vety, preto naznačíme len záchytné body dôkazu. Označme $g(x) = x - p\frac{f(x)}{f'(x)}$. Zrejme $f(x) = (x-\bar{x})^p h(x)$ pričom $h(\bar{x}) \neq 0$ a jednoduchým výpočtom dostaneme

$$\begin{aligned} f'(x) &= (x-\bar{x})^p h'(x) + p(x-\bar{x})^{p-1}h(x), \\ g(x) &= x - p\frac{(x-\bar{x})^p h(x)}{(x-\bar{x})^p h'(x) + p(x-\bar{x})^{p-1}h(x)} \\ &= x - p\frac{(x-\bar{x})h(x)}{(x-\bar{x})h'(x) + ph(x)} \\ &= x - p\frac{C(x)}{D(x)}. \end{aligned}$$

Formálne môžeme písať

$$\begin{aligned} g'(x) &= 1 - p\frac{C'(x)D(x) - C(x)D'(x)}{D^2(x)}, \\ g'(\bar{x}) &= 1 - p\frac{h^2(\bar{x})p}{h^2(\bar{x})p^2} = 0. \end{aligned}$$ □

Poznámka 6.3 *Prakticky odhadujeme chybu aproximácie vypočítanej pomocou Newtonovej metódy takto. Keďže*

$$f(x_n) = f(x_n) - f(\bar{x}) = f'(\xi_n)(x_n - \bar{x}),$$

platí

$$\bar{x} - x_n = \frac{-f(x_n)}{f'(\xi_n)} \approx \frac{-f(x_n)}{f'(x_n)} = x_{n+1} - x_n,$$

t.j. $\bar{x} - x_n \approx x_{n+1} - x_n$.

Pre relatívnu chybu používame odhad

$$\frac{\bar{x} - x_n}{\bar{x}} \approx \frac{x_{n+1} - x_n}{x_{n+1}}.$$

□

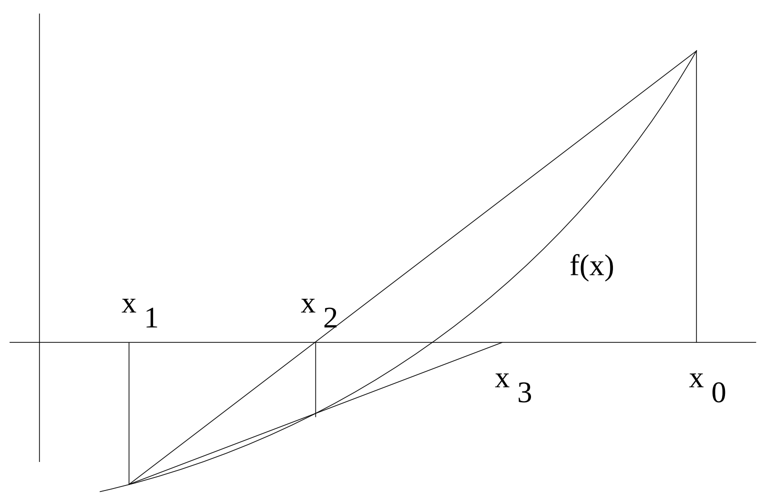
ALGORITMUS NEWTON

Vstup $f, f', x_0, \varepsilon, imax, ier$ (identifikátor chyby a ukončenia)

1. $i = 1$
2. $men = f'(x_0)$
3. ak $men = 0 \Rightarrow ier = 2$ Koniec
4. $x_1 = x_0 - f(x_0)/men$
5. ak $|x_1 - x_0| \leq \varepsilon \quad ier = 1$ Koniec
6. ak $i > imax \Rightarrow ier = 3$ Koniec

   ináč $i = i + 1, x_0 = x_1$ choď na 2.

Metóda sečníc je modifikácia Newtonovej metódy, kde $f'(x_n)$ nahradíme pomocou $\frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}$, t.j. $x_{n+1} = x_n - \frac{f(x_n)}{\frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}}$. Geometricky to znamená, že dotyčnicu ku grafu $f(x)$ nahradíme sečnicou (pozri obrázok 6.5).



Obr. 6.5. Metóda sečníc

Zjednodušená Newtonova metóda je daná

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)},$$

t.j. $f'(x_n)$ sme nahradili $f'(x_0)$. Je to vlastne metóda $x_{n+1} = x_n + Mf(x_n)$, kde $M = -1/f'(x_0)$.

Nasledujúca metóda umožňuje získať dvojstranný odhad koreňa, a teda aj chybu aproximácie.

Veta 6.8 (Newtonova-Fourierova metóda) *Nech funkcia* $f \in C^2([a,b])$ *je taká, že* $f(a) < 0 < f(b)$ *a pre všetky* $x \in [a,b]$ *platí* $f'(x) > 0, f''(x) > 0$. *Ďalej nech* $z_0 = a, x_0 = b$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad n = 0, 1, \ldots, \tag{6.13}$$

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(x_n)} \quad n = 0, 1, \ldots \ . \tag{6.14}$$

Potom

i) existuje jediné $\bar{x} \in [a,b]$ *také, že* $f(\bar{x}) = 0$,

ii) $x_n \to \bar{x}, \quad x_{n+1} < x_n,$
$z_n \to \bar{x}, \quad z_{n+1} > z_n,$

iii) $\lim_{n\to\infty} \dfrac{x_{n+1} - z_{n+1}}{(x_n - z_n)^2}$ *je konečná.*

DÔKAZ: *i)* Spojitosť funkcie f a $f(a) < 0 < f(b)$ implikuje existenciu $\bar{x} \in (a,b)$, pričom $f(\bar{x}) = 0$. Striktná rastúcosť f hovorí, že $\bar{x}$ je jednoznačne určené.

ii) Najprv dokážeme $z_0 < z_1 < \bar{x} < x_1 < x_0$. Zrejme

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{f(b)}{f'(b)} < x_0,$$

$$z_1 = z_0 - \frac{f(z_0)}{f'(x_0)} = z_0 - \frac{f(a)}{f'(b)} > z_0.$$

Keďže

$$0 = f(\bar{x}) = f(x_n) + f'(x_n)(\bar{x} - x_n) + \frac{1}{2} f''(\xi_n)(\bar{x} - x_n)^2$$

platí pre nejaké ξ_n medzi $\bar{x}, x_n$, po úprave dostaneme

$$\bar{x} - x_{n+1} = \frac{-(\bar{x} - x_n)^2}{2} \frac{f''(\xi_n)}{f'(x_n)}, \tag{6.15}$$

a preto $\bar{x} - x_{n+1} < 0$, čiže $x_{n+1} > \bar{x}$.

Ďalej z vety o strednej hodnote vyplýva

$$\begin{aligned}
\bar{x} - z_1 &= \bar{x} - z_0 + \frac{f(z_0)}{f'(x_0)} = \bar{x} - z_0 + \frac{f(z_0) - f(\bar{x})}{f'(x_0)} \\
&= \bar{x} - z_0 + f'(\xi_0) \frac{(z_0 - \bar{x})}{f'(x_0)} \\
&= (\bar{x} - z_0) \frac{f'(x_0) - f'(\xi_0)}{f'(x_0)} > 0,
\end{aligned}$$

lebo f' je rastúca. Platí teda

$$z_0 < z_1 < \bar{x} < x_1 < x_0.$$

Indukciou dostaneme

$$z_0 < z_1 < \ldots < z_n < \bar{x} < x_n < \ldots < x_1 < x_0.$$

Preto

$$x_n \to \alpha \geq \bar{x},$$
$$z_n \to \beta \leq \bar{x}.$$

Limitným prechodom v (6.13) a (6.14) dostávame $f(\alpha) = f(\beta) = 0$, teda platí *ii*).

iii) Postupne použijeme dvakrát vetu o strednej hodnote a pre nejaké $\xi_n \in (z_n, x_n)$ i $\theta_n \in (\xi_n, x_n)$ platí

$$\begin{aligned} x_{n+1} - z_{n+1} &= x_n - z_n + \frac{f(z_n) - f(x_n)}{f'(x_n)} && = x_n - z_n + \frac{f'(\xi_n)}{f'(x_n)}(z_n - x_n) \\ &= (x_n - z_n)\frac{f'(x_n) - f'(\xi_n)}{f'(x_n)} && = (x_n - z_n)(x_n - \xi_n)\frac{f''(\theta_n)}{f'(x_n)} \\ &\leq (x_n - z_n)^2 \frac{f''(\theta_n)}{f'(x_n)}. \end{aligned}$$

Teda

$$0 \leq \frac{x_{n+1} - z_{n+1}}{(x_n - z_n)^2} \leq \frac{f''(\theta_n)}{f'(x_n)} \leq C,$$

lebo $\theta_n, x_n \to \bar{x}$. □

Príklad 6.5 Uvažujme reálnu funkciu $f : [0,3; 0,8] \mapsto \mathbb{R}$ definovanú predpisom $f(x) = x - \cos x$. Pomocou odhadu chyby pre Newtonovu metódu určte čo najmenší počet Newtonových iterácií na výpočet riešenia $f(x) = 0$, ak chyba aproximácie má byť menšia ako 10^{-8}.

Riešenie: Derivovaním dostaneme $f'(x) = 1 + \sin x$, $f''(x) = \cos x$. Na intervale $[0,3; 0,8]$ sú splnené predpoklady vety 6.8 a platí

$$\left|\frac{f''(y)}{f'(z)}\right| \leq 1, \qquad \forall y, z \in [0,3; 0,8].$$

Použijeme vzťah (6.15) a dostaneme

$$|\bar{x} - x_{n+1}| \quad = \left|\frac{(\bar{x} - x_n)^2}{2} \frac{f''(\xi_n)}{f'(x_n)}\right| \leq (\bar{x} - x_n)^2 \quad \leq \ldots \leq (\bar{x} - x_1)^{2^n} \leq \left(\frac{1}{2}\right)^{2^n}.$$

Povolená chyba je 10^{-8}, teda

$$\left(\frac{1}{2}\right)^{2^n} < 10^{-8}$$

je splnené napr. pre $n = 5$. K tomuto výsledku sa dá dospieť logaritmovaním posledného vzťahu. □

6.3 Urýchľovanie konvergencie, Aitkenov Δ^2 proces

Nech x_n je postupnosť zostrojená pomocou

$$x_{n+1} = g(x_n).$$

Nech $0 < g'(\xi) < 1$ na $[a,b]$. Vieme, že vo všeobecnosti

$$x_{n+1}-\bar{x} = g(x_n) - g(\bar{x}) = g'(\xi_n)(x_n - \bar{x}),$$

t.j.

$$\begin{aligned} \frac{x_{n+1}-\bar{x}}{x_n-\bar{x}} &= g'(\xi_n), \\ \lim_{n\to\infty} \frac{x_{n+1}-\bar{x}}{x_n-\bar{x}} &= g'(\bar{x}) = \lim_{n\to\infty} \frac{x_n-\bar{x}}{x_{n-1}-\bar{x}}. \end{aligned}$$

Tvrdenie: $\lambda_n := \frac{x_n - x_{n-1}}{x_{n-1}-x_{n-2}} \to g'(\bar{x})$ pre $n \to \infty$.
Z výpočtu dostaneme

$$\begin{aligned} \lambda_n &= \frac{\bar{x}-x_{n-1}-(\bar{x}-x_n)}{\bar{x}-x_{n-2}-(\bar{x}-x_{n-1})} &&= \frac{\bar{x}-x_{n-1}-g'(\xi_{n-1})(\bar{x}-x_{n-1})}{(\bar{x}-x_{n-1})/g'(\xi_{n-2})-(\bar{x}-x_{n-1})} \\ &= \frac{1-g'(\xi_{n-1})}{1/g'(\xi_{n-2})-1} &&\to \frac{1-g'(\bar{x})}{\frac{1}{g'(\bar{x})}-1} = g'(\bar{x}). \end{aligned}$$

Pre veľké n teda $\lambda_n \approx g'(\bar{x})$, a preto $\frac{\bar{x}-x_n}{\bar{x}-x_{n-1}} \approx \lambda_n$. Odtiaľ $\bar{x}-x_n \approx \lambda_n(\bar{x}-x_{n-1})$, a teda

$$\bar{x}-x_n = (\bar{x}-x_{n-1}) + (x_{n-1}-x_n) \approx \frac{1}{\lambda_n}(\bar{x}-x_n) + (x_{n-1}-x_n),$$

čiže

$$(\bar{x}-x_n)\left(1-\frac{1}{\lambda_n}\right) \approx x_{n-1}-x_n.$$

Odtiaľ vyplýva Aitkenova chybová formula

$$\bar{x}-x_n \approx \frac{\lambda_n}{1-\lambda_n}(x_n-x_{n-1})$$

a Aitkenova extrapolačná formula

$$\bar{x} \approx x_n + \frac{\lambda_n}{1-\lambda_n}(x_n-x_{n-1}) =: \hat{x}_n.$$

Aitkenova extrapolačná formula sa píše obyčajne v tvare

$$\hat{x}_n = x_n - \frac{(x_n-x_{n-1})^2}{(x_n-x_{n-1})-(x_{n-1}-x_{n-2})} = x_n - \frac{(\Delta x_n)^2}{\Delta^2 x_n}. \tag{6.16}$$

Postupnosť $\hat{x}_n$ obyčajne konverguje rýchlejšie než x_n.

ALGORITMUS

Vstup: x_0, g, ε
1. $x_1 = g(x_0), x_2 = g(x_1)$
2. $\hat{x}_2 = x_2 - \frac{(x_2-x_1)^2}{(x_2-x_1)-(x_1-x_0)}$
3. ak $|\hat{x}_2 - x_2| \le \varepsilon \Rightarrow \hat{x}_2$ je koreň, Koniec
   ináč $x_0 = \hat{x}_2$ choď na 1.

6.4 Metóda bisekcie

Je založená na tomto postupe: Ak spojitá funkcia f spĺňa $f(a)f(b) < 0$, tak bude mať medzi a, b koreň. Určí sa hodnota $f((a+b)/2)$. Ak $f((a+b)/2) \neq 0$, tak podľa znamienka $f((a+b)/2)$ sa zistí, či bude koreň ležať medzi $a, (a+b)/2$ alebo $b, (a+b)/2$. Postup sa opakuje, kým sa neurčí približná hodnota koreňa $f(x) = 0$ s dostatočnou presnosťou. Metóda bisekcie sa niekedy nazýva metódou polenia intervalu.

Veta 6.9 *Nech $f \in C([a,b]), f(a)f(b) < 0$. Potom postupnosť x_n zostrojená metódou bisekcie konverguje k $\bar{x} \in [a,b]$ pričom $f(\bar{x}) = 0$. Platí*

$$|x_n - \bar{x}| \leq \frac{b-a}{2^n}.$$

Dôkaz ponechávame na čitateľa. □

ALGORITMUS METÓDY BISEKCIE

<u>Vstup:</u> f, a, b, ε

1. $c = \frac{1}{2}(a+b)$
2. ak $(b-c) \leq \varepsilon \Rightarrow$ koreň je c Koniec
3. ak $f(b)f(c) \leq 0 \Rightarrow a = c$
   ináč $b = c$
4. choď na 1.

6.5 Viacrozmerný Newtonov algoritmus

Vety o pevnom bode platia aj vo viacrozmernom prípade, t.j. v $\mathbb{R}^n$. Pre jednoduchosť si uveďme len ich znenia bez dôkazov.

Veta 6.10 (Banachova) *Nech (X,d) je úplný metrický priestor a nech pre zobrazenie $f : X \to X$ platí*

$$d(f(x), f(y)) \leq \lambda d(x,y), \qquad \forall x, y \in X$$

pre nejaké λ s vlastnosťou $0 \leq \lambda < 1$. Potom existuje práve jeden bod $x^* \in X$, pre ktorý

$$f(x^*) = x^*.$$

Navyše pre ľubovoľný bod $x_0 \in X$ platí

$$x_{n+1} := f(x_n) \longrightarrow x^*, \qquad \text{pre } n \to \infty.$$

Veta 6.11 (Brouwerova) *Nech $n \in \mathbb{N}$, $M \subset \mathbb{R}^n$ je kompaktná a konvexná množina. Ak funkcia $f : M \longrightarrow M$ je spojitá, tak má pevný bod, t.j. existuje také $\boldsymbol{x}^* \in M$, že $f(\boldsymbol{x}^*) = \boldsymbol{x}^*$.*

Iteračné metódy pre riešenie $f(x) = 0$, ktoré sme si doteraz opísali, sa dajú zovšeobecniť i pre viacrozmernú úlohu. V tejto časti si odvodíme vzorce pre Newtonovu aproximačnú metódu vo viacrozmernom prípade. Pre jednoduchosť sa obmedzíme na dvojrozmerný prípad, lebo tam sa dajú matice ľahko invertovať.

Uvažujme rovnice

$$f(x,y) = 0, \qquad g(x,y) = 0. \tag{6.17}$$

Nech (s,t) je ich riešením. Preveďme linearizáciu takto

$$s = x + \xi, \qquad t = y + \eta,$$

teda dostaneme

$$\begin{aligned} 0 &= f(s,t) &&= f(x+\xi, y+\eta) &&\approx f(x,y) + \xi f_x(x,y) + \eta f_y(x,y), \\ 0 &= g(s,t) &&= g(x+\xi, y+\eta) &&\approx g(x,y) + \xi g_x(x,y) + \eta g_y(x,y), \end{aligned}$$

kde f_x, f_y, g_x, g_y sú príslušné parciálne derivácie, napr. $f_x = \frac{\partial f}{\partial x}$. Označme

$$\boldsymbol{F} = \begin{pmatrix} f \\ g \end{pmatrix}, \qquad \boldsymbol{\Phi} = \begin{pmatrix} f_x & f_y \\ g_x & g_y \end{pmatrix}.$$

Môžeme písať'

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \approx \boldsymbol{F}(x,y) + \boldsymbol{\Phi}(x,y) \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

Zaveďme nasledujúce iterácie

$$\xi = x_{k+1} - x_k, \qquad \eta = y_{k+1} - y_k,$$

teda

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \boldsymbol{F}(x_k, y_k) + \boldsymbol{\Phi}(x_k, y_k) \left[\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} - \begin{pmatrix} x_k \\ y_k \end{pmatrix} \right].$$

Z jednoduchého výpočtu vyplýva

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \boldsymbol{\Phi}^{-1}(x_k, y_k) \boldsymbol{F}(x_k, y_k).$$

V dvojrozmernom prípade sa dá inverzná matica 2×2 ľahko vypočítať' , t.j.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Takže napokon dostaneme

$$\begin{aligned} x_{k+1} &= x_k + \frac{f_y(x_k,y_k) g(x_k,y_k) - g_y(x_k,y_k) f(x_k,y_k)}{f_x(x_k,y_k) g_y(x_k,y_k) - f_y(x_k,y_k) g_x(x_k,y_k)} \\ &= x_k + \left. \frac{f_y g - g_y f}{f_x g_y - f_y g_x} \right|_{(x_k,y_k)}, \\ y_{k+1} &= y_k + \frac{g_x(x_k,y_k) f(x_k,y_k) - g(x_k,y_k) f_x(x_k,y_k)}{f_x(x_k,y_k) g_y(x_k,y_k) - f_y(x_k,y_k) g_x(x_k,y_k)} \\ &= y_k + \left. \frac{g_x f - g f_x}{f_x g_y - f_y g_x} \right|_{(x_k,y_k)}. \end{aligned} \tag{6.18}$$

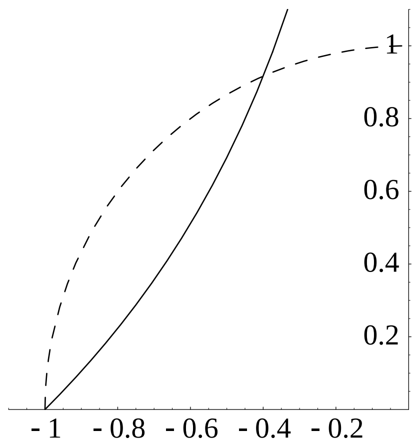
Poznámka 6.4 *Vo viacrozmernom prípade príslušný systém lineárnych rovníc neriešime hľadaním inverznej matice (ako to bolo vo vyššie uvedenom dvojrozmernom prípade), ale použitím niektorej z metód na riešenie systémov lineárnych rovníc.*

Príklad 6.6 Dané sú rovnice

$$\begin{aligned} f(x,y) &= 1 + xe^y &&= 0, \\ g(x,y) &= x^2 + y^2 - 1 &&= 0. \end{aligned}$$

Nájdite vhodné štartovacie hodnoty pre x_0, y_0 a vypočítajte prvú aproximáciu riešenia pomocou Newtonovej metódy.

Riešenie: Nakreslime si najprv grafy riešení jednotlivých rovníc (viď obrázok 6.6). Za vhodné štartovacie hodnoty pre Newtonov algoritmus môžeme zvoliť $x_0 = -e^{-1}$, $y_0 = 1$.



Obr. 6.6. Náčrt pre získanie x_0, y_0. Riešenie $f(x,y) = 0$ je znázornené plnou a $g(x,y) = 0$ čiarkovanou čiarou

Pre derivácie funkcií f, g platí

$$\begin{aligned} f_x &= e^y, & g_x &= 2x, \\ f_y &= xe^y, & g_y &= 2y. \end{aligned}$$

Dosadením do (6.18) dostaneme

$$\begin{aligned} x_{k+1} &= x_k + \frac{x_k e^{y_k}(x_k^2 + y_k^2 - 1) - 2y_k\,(1 + x_k e^{y_k})}{e^{y_k} 2y_k - x_k e^{y_k} 2x_k}, \\ y_{k+1} &= y_k + \frac{2x_k\,(1 + x_k e^{y_k}) - (x_k^2 + y_k^2 - 1)e^{y_k}}{e^{y_k} 2y_k - x_k e^{y_k} 2x_k}. \end{aligned}$$

Takže pre prvú aproximáciu máme

$$x_1 \approx -0,371776 \qquad y_1 \approx 0,921741\ .$$

□

Príklad 6.7 Dané sú rovnice

$$\begin{aligned} f(x,y) &= 1 + xe^y &&= 0, \\ g(x,y) &= x^2 + y^2 - 1 &&= 0. \end{aligned}$$

Transformujte 2-rozmernú úlohu na 1-rozmernú pomocou eliminácie jednej neznámej. Určte vhodné štartovacie hodnoty pre x_0, y_0 a vypočítajte prvé štyri aproximácie riešenia pomocou Newtonovej metódy.

Riešenie: Pretože $1 + xe^y = 0 \Rightarrow x = -e^{-y}$, dosadením do druhej rovnice dostaneme

$$0 = x^2 + y^2 - 1 = e^{-2y} + y^2 - 1 = h(y).$$

Pre deriváciu máme $h'(y) = -2e^{-2y} + 2y > 0$ na intervale $[0,5;1]$. Pre druhú deriváciu platí $h''(y) = 4e^{-2y} + 2 > 0$. Ďalej vieme, že $h(0,5) \approx -0,38 < 0$ a $h(1) = e^{-2} > 0$. Takže funkcia h bude mať koreň v intervale $[0,5;1]$. Iteračný predpis znie

$$y_{k+1} = y_k - \frac{h(y_k)}{h'(y_k)}, \qquad x_k = -e^{-y_k}.$$

Výpočtom dostaneme

k	0	1	2	3	4
x_k	-0,367879	-0,397826	-0,399882	-0,399881	-0,399881
y_k	1	0,921741	0,916586	0,916563	0,916563

.

□

Cvičenie 6.1 Nakreslite graficky rôzne prípady $x = g(x)$ zo vzťahu (6.8) a demonštrujte graficky postupnosť $x_{n+1} = g(x_n)$.

Cvičenie 6.2 Nech $p > 1$. Akú hodnotu nadobúda nekonečný zlomok

$$\cfrac{1}{p + \cfrac{1}{p + \cfrac{1}{p + \dots}}} \; ?$$

Matematicky to zdôvodnite.

Cvičenie 6.3 Podmienku $\lambda < 1$ v Banachovej vete o pevnom bode nemôžeme vynechať. Ukážte, že funkcia

$$f(x) = \begin{cases} x - \frac{e^x}{2} & \text{pre } x \le 0 \\ -\frac{1}{2} + \frac{1}{2}x & \text{pre } x \ge 0 \end{cases}$$

nemá žiadny pevný bod, hoci $|f(x) - f(y)| < |x - y|$ pre všetky $x, y \in \mathbb{R}$.

Cvičenie 6.4 Vezmite si kalkulačku. Zvoľte na nej ľubovoľné reálne číslo. Akú hodnotu získate, ak na nej dostatočne veľakrát stlačíte tlačítko "COS"? Vašu odpoveď matematicky zdôvodnite.

Cvičenie 6.5 Napíšte algoritmus pre metódu sečníc.

Cvičenie 6.6 Dokážte, že platí (6.16).

Cvičenie 6.7 Použite Newtonovu metódu na odvodenie postupnosti, ktorá konverguje k číslu $2^{\frac{1}{3}}$. Vzorec pre výpočet členov postupnosti smie obsahovať iba aritmetické operácie +, -, . ,: . Použite štartovaciu hodnotu $x_0 = 1$ a vypočítajte dve iterácie.

Cvičenie 6.8 Daná je konvexná funkcia f na intervale $[a, b]$, pričom platí nerovnosť $f(a) < 0 < f(b)$. Aproximácie riešenia rovnice $f(x) = 0$ zvoľme nasledujúcim spôsobom[1]. Začneme v bode $x_0 = a$. Postupne pre $k = 1, 2, \dots$ získame x_k ako prienik priamky (tetivy) danej bodmi $(x_{k-1}, f(x_{k-1})), (b, f(b))$ so súradnicovou osou x. Nakreslite si obrázok a napíšte vzorec na výpočet postupnosti aproximácií $\{x_k\}_{k=1}^{\infty}$.

Cvičenie 6.9 Modifikujte aproximačnú metódu z predošlého cvičenia pre prípad konkávnej funkcie[2]. Ako treba zmeniť body pri voľbe tetivy?

[1] Táto metóda sa nazýva *metóda Regula falsi* pre konvexné funkcie.

[2] Takto získaná metóda sa nazýva *metóda Regula falsi* pre konkávne funkcie.

Kapitola 7

Riešenie systému lineárnych algebraických rovníc iteračnými metódami

Numerické riešenie parciálnych diferenciálnych rovníc s danými okrajovými resp. i začiatočnými podmienkami vedie často k riešeniu veľkých systémov lineárnych algebraických rovníc. Nezáleží na tom, akú numerickú metódu (konečné prvky, konečné diferencie, alebo iné) sme na riešenie diferenciálnych rovníc pritom použili. Takto získaný systém lineárnych algebraických rovníc má však špeciálnu štruktúru. Jeho matica je síce veľká, no riedka, t.j. len malá časť jej prvkov je nenulová. Okrem toho matica môže byť symetrická. V podstate sa riešenie takéhoto systému rovníc dá dostať priamymi metódami, napr. Gaussovou eliminačnou metódou. V priebehu eliminácie však počet nenulových elementov matice prudko narastá. Týmto môžu nastať problémy s počítačovou pamäťou. Preto iteračné metódy predstavujú vhodnú alternatívu na riešenie systémov s veľkými no riedkymi maticami.

Normy
Úvodom si pripomeňme niektoré základné pojmy z lineárnej algebry, ktoré budeme v ďalšom texte využívať.

Definícia 7.1 *Vektorovou normou $\|\boldsymbol{x}\|$ vektora $\boldsymbol{x} = (x^1, \ldots, x^n)^T \in \mathbb{R}^n$ nazývame reálnu funkciu jeho zložiek, pričom*

i) $\|\boldsymbol{x}\| \geq 0,\ \|\boldsymbol{x}\| = 0 \Longleftrightarrow \boldsymbol{x} = \boldsymbol{0}$,

ii) $\|c\boldsymbol{x}\| = |c|\ \|\boldsymbol{x}\|,\ \forall c \in \mathbb{R},\ \boldsymbol{x} \in \mathbb{R}^n$,

iii) $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ *pre všetky* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

Uveďme pre názornosť niektoré dôležité príklady vektorových noriem:

Maximová norma: $\|\boldsymbol{x}\|_\infty = \max\limits_{1 \leq k \leq n} |x^k|$,

l_p-norma: $\|\boldsymbol{x}\|_p = \left(\sum\limits_{k=1}^{n} |x^k|^p \right)^{\frac{1}{p}}$ pre $p \geq 1$.

Poznamenajme, že l_2-norma sa nazýva *euklidovskou* normou.

Definícia 7.2 *Normou matice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ rozumieme reálnu funkciu $\|\boldsymbol{A}\|$ prvkov matice s vlastnosťami* *i)* $\|\boldsymbol{A}\| \geq 0$ *pre všetky* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, *pričom* $\|\boldsymbol{A}\| = 0$ *len pre* $\boldsymbol{A} = \boldsymbol{0}$, *t.j. matica* A *je nulová matica,*

ii) $\|c\boldsymbol{A}\| = |c|\ \|\boldsymbol{A}\|$ *pre všetky* $c \in \mathbb{R}$ *a* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$,

iii) $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$ *pre ľubovoľné matice* $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$,

iv) $\|\boldsymbol{A} \cdot \boldsymbol{B}\| \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|$ *pre ľubovoľné matice* $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$.

Často používanými príkladmi maticových noriem sú

Maximová norma: $\|\boldsymbol{A}\|_M = n \max\limits_{1\leq i,j\leq n} |a_{ij}|$,

Riadková norma: $\|\boldsymbol{A}\|_R = \max\limits_{1\leq i\leq n} \sum\limits_{j=1}^{n} |a_{ij}|$,

Stĺpcová norma: $\|\boldsymbol{A}\|_S = \max\limits_{1\leq j\leq n} \sum\limits_{i=1}^{n} |a_{ij}|$,

Frobeniova norma: $\|\boldsymbol{A}\|_F = \sqrt{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} |a_{ij}|^2}$,

Spektrálna norma: $\|\boldsymbol{A}\|_{Sp} = \max\limits_{i} \sqrt{\lambda_i\left(\boldsymbol{A}^T.\boldsymbol{A}\right)}$, kde symbol $\lambda_i\left(\boldsymbol{A}^T.\boldsymbol{A}\right)$ označuje vlastné číslo matice $\boldsymbol{A}^T.\boldsymbol{A}$.

Definícia 7.3 *Hovoríme, že maticová norma je* súhlasná *resp.* kompatibilná *s vektorovou normou, ak nezávisle od voľby matice i vektora platí*

$$\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|\ \|\boldsymbol{x}\|.$$

Nasledujúce kombinácie maticových a vektorových noriem sú kompatibilné:

- $\|\boldsymbol{A}\|_M$ alebo $\|\boldsymbol{A}\|_R$ sú súhlasné s $\|\boldsymbol{x}\|_\infty$,
- $\|\boldsymbol{A}\|_M$ alebo $\|\boldsymbol{A}\|_S$ sú súhlasné s $\|\boldsymbol{x}\|_1$,
- $\|\boldsymbol{A}\|_\infty$ alebo $\|\boldsymbol{A}\|_F$ sú súhlasné s $\|\boldsymbol{x}\|_2$.

V ďalšom texte budeme predpokladať, že maticová i vektorová norma budú navzájom kompatibilné.

7.1 Metóda postupných aproximácií

K najjednoduchším metódam patrí *metóda postupných aproximácií*, ktorú si teraz objasníme.

Uvažujme systém lineárnych rovníc tvaru

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{7.1}$$

kde $\boldsymbol{A}$ je regulárna štvorcová matica a vektor $\boldsymbol{b}$ je pravá strana systému. Označme $\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{A}$ a $\boldsymbol{g} = \boldsymbol{b}$. Systém (7.1) prepíšme do tvaru

$$\boldsymbol{x} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{g}. \tag{7.2}$$

Postupne počítajme aproximácie podľa predpisu

$$\boldsymbol{x}_k = \boldsymbol{C}\boldsymbol{x}_{k-1} + \boldsymbol{g}. \tag{7.3}$$

Tento postup sa nazýva metóda postupných aproximácií. Ak daný iteračný postup konverguje, t.j. $\lim\limits_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^*$, tak limitným prechodom v (7.3) zistíme, že $\boldsymbol{x}^*$ je riešením (7.2), a teda i (7.1). O konvergencii metódy hovorí nasledujúca veta.

Veta 7.1 *Nech všetky vlastné čísla matice $\boldsymbol{C}$ sú v absolútnej hodnote menšie než 1. Potom metóda postupných aproximácií konverguje pre ľubovoľnú voľbu $\boldsymbol{x}_0$.*

DÔKAZ: Pomocou matematickej indukcie dostaneme

$$\boldsymbol{x}_k = \boldsymbol{C}^k\boldsymbol{x}_0 + (\boldsymbol{I} + \boldsymbol{C} + \ldots + \boldsymbol{C}^{k-1})\boldsymbol{g}. \tag{7.4}$$

Platí

$$(\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{C}^2 + \ldots + \boldsymbol{C}^{k-1})(\boldsymbol{I} - \boldsymbol{C}) = \boldsymbol{I} - \boldsymbol{C}^k.$$

Vynásobením sprava maticou $(\boldsymbol{I} - \boldsymbol{C})^{-1}$ máme

$$\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{C}^2 + \ldots + \boldsymbol{C}^{k-1} = (\boldsymbol{I} - \boldsymbol{C})^{-1} - \boldsymbol{C}^k(\boldsymbol{I} - \boldsymbol{C})^{-1}.$$

Ak sú všetky vlastné čísla matice $\boldsymbol{C}$ v absolútnej hodnote menšie než 1, potom $\boldsymbol{C}^k \stackrel{k\to\infty}{\longrightarrow} \boldsymbol{0}$ a

$$(\boldsymbol{I} + \boldsymbol{C} + \ldots + \boldsymbol{C}^{k-1}) \stackrel{k\to\infty}{\longrightarrow} (\boldsymbol{I} - \boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} \ ,$$

t.j. $\sum\limits_{k=0}^{\infty} \boldsymbol{C}^k = (\boldsymbol{I} - \boldsymbol{C})^{-1}$. □

Veta 7.2 *Nech $\|C\|$ je norma matice C a platí $\|\boldsymbol{C}\| < 1$, potom metóda postupných aproximácií konverguje pre ľubovoľnú voľbu $\boldsymbol{x}_0$, $\lim\limits_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^*$ a $\boldsymbol{x}^*$ je riešenie (7.1). Pre odhad chyby platí*

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \le \|\boldsymbol{C}\|^k\|\boldsymbol{x}_0\| + \frac{\|\boldsymbol{g}\| \ \|\boldsymbol{C}\|^k}{1 - \|\boldsymbol{C}\|}.$$

DÔKAZ: Zo vzťahu $\sum\limits_{k=0}^{\infty} \boldsymbol{C}^k = (\boldsymbol{I} - \boldsymbol{C})^{-1}$ vyplýva

$$(\boldsymbol{I} - \boldsymbol{C})^{-1} - (\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{C}^2 + \ldots + \boldsymbol{C}^k) = \sum_{i=k+1}^{\infty} \boldsymbol{C}^i,$$

teda

$$\begin{aligned}
\|(\boldsymbol{I} - \boldsymbol{C})^{-1} - (\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{C}^2 + \ldots + \boldsymbol{C}^k)\| &\le \sum_{i=k+1}^{\infty} \|\boldsymbol{C}\|^i \\
&= \|\boldsymbol{C}\|^{k+1} \sum_{i=0}^{\infty} \|\boldsymbol{C}\|^i \\
&= \frac{\|\boldsymbol{C}\|^{k+1}}{1 - \|\boldsymbol{C}\|}.
\end{aligned}$$

Pomocou (7.4) dostaneme

$$\begin{aligned}
\|\boldsymbol{x}^* - \boldsymbol{x}_k\| &= \|(\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{g} - (\boldsymbol{I}+\boldsymbol{C}+\ldots+\boldsymbol{C}^{k-1})\boldsymbol{g} - \boldsymbol{C}^k\boldsymbol{x}_0\| \\
&\le \|(\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{g} - (\boldsymbol{I}+\boldsymbol{C}+\ldots+\boldsymbol{C}^{k-1})\boldsymbol{g}\| + \|\boldsymbol{C}^k\boldsymbol{x}_0\| \\
&\le \|(\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{g} - (\boldsymbol{I}+\boldsymbol{C}+\ldots+\boldsymbol{C}^{k-1})\boldsymbol{g}\| + \|\boldsymbol{C}\|^k\|\boldsymbol{x}_0\| \\
&\le \frac{\|\boldsymbol{g}\|\ \|\boldsymbol{C}\|^k}{1-\|\boldsymbol{C}\|} + \|\boldsymbol{C}\|^k\|\boldsymbol{x}_0\|.
\end{aligned}$$

Odtiaľ vyplýva tvrdenie vety. □

O porovnaní presnosti dvoch po sebe idúcich iterácií hovorí ďalšia veta.

Veta 7.3 *Pre ľubovoľné k platí*

$$\|\boldsymbol{x}^* - \boldsymbol{x}_k\| \le \|\boldsymbol{C}\|\ \|\boldsymbol{x}^* - \boldsymbol{x}_{k-1}\|.$$

DÔKAZ: Zrejme

$$\|\boldsymbol{x}^* - \boldsymbol{x}_k\| = \|(\boldsymbol{C}\boldsymbol{x}^* + \boldsymbol{g}) - (\boldsymbol{C}\boldsymbol{x}_{k-1} + \boldsymbol{g})\| \le \|\boldsymbol{C}\|\ \|\boldsymbol{x}^* - \boldsymbol{x}_{k-1}\|.$$ □

7.2 Jacobiho iteračná metóda

Jacobiho metóda predstavuje istú modifikáciu metódy postupných aproximácií. Objasnime si, v čom spočíva.

Nech $\boldsymbol{A}$ je regulárna matica rádu $n \times n$. Predpokladajme, že jej diagonálne prvky sú nenulové. Chceme riešiť systém lineárnych rovníc (7.1) pre danú pravú stranu $\boldsymbol{b} \in \mathbb{R}^n$. Rozložme si maticu $\boldsymbol{A}$ na tri časti: $\boldsymbol{L}$ (dolnú trojuholníkovú maticu), $\boldsymbol{D}$ (diagonálu), $\boldsymbol{U}$ (hornú trojuholníkovú maticu), pričom platí $\boldsymbol{A} = -\boldsymbol{L} + \boldsymbol{D} - \boldsymbol{U}$. Teda

$$\boldsymbol{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} a_{1,1} & 0 & 0 & \ldots & 0 \\ 0 & a_{2,2} & 0 & \ldots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & a_{n,n} \end{pmatrix},$$

$$\boldsymbol{L} = -\begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ a_{2,1} & 0 & 0 & \ldots & \vdots \\ a_{3,1} & a_{3,2} & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ a_{n,1} & \ldots & \ldots & a_{n,n-1} & 0 \end{pmatrix}, \quad \boldsymbol{U} = -\begin{pmatrix} 0 & a_{1,2} & a_{1,3} & \ldots & a_{1,n} \\ 0 & 0 & a_{2,3} & \ldots & a_{2,n} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \ldots & \ldots & 0 & 0 \end{pmatrix}.$$

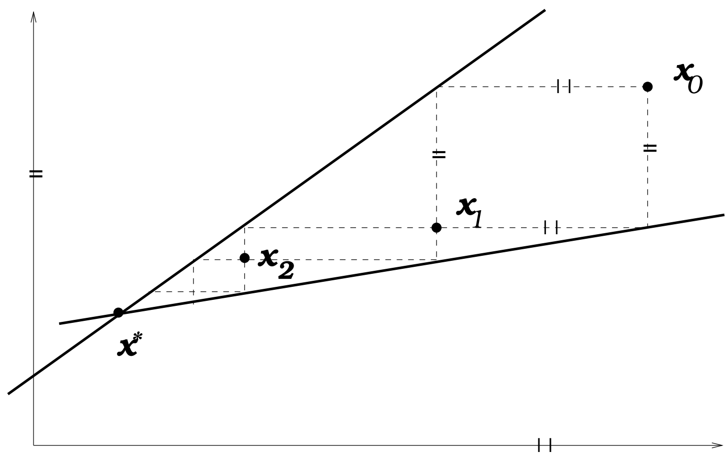
Systém (7.1) prepíšeme do tvaru

$$\boldsymbol{x} = \boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U})\boldsymbol{x} + \boldsymbol{D}^{-1}\boldsymbol{b} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}^{-1}\boldsymbol{b},$$

pričom $\boldsymbol{C} = \boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U})$. Táto matica je iteračnou maticou pre Jacobiho metódu, ktorá je daná predpisom

$$\boldsymbol{x}_{k+1} = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{D}^{-1}\boldsymbol{b}. \tag{7.5}$$

Pre názornosť sme Jacobiho iteračnú metódu pre dve rovnice graficky zobrazili na obrázku 7.1.



Obr. 7.1. Jacobiho metóda v $\mathbb{R}^2$

O jej konvergencii hovorí nasledujúca veta.

Veta 7.4 *Nech $\boldsymbol{A}$ je regulárna matica rádu $n \times n$, pričom jej diagonálne prvky sú nenulové. Ak $\|\boldsymbol{C}\| = \|\boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U})\| < 1$, potom Jacobiho iteračná metóda daná predpisom (7.5) konverguje pre ľubovoľný štartovací vektor $\boldsymbol{x}_0$ ku riešeniu (7.1).*

DÔKAZ: Stačí použiť vetu 7.2. □

Dôsledok 7.1 *Ak je splnené niektoré z týchto kritérií:*

(i) riadkové kritérium

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{i,j}| < |a_{i,i}|, \qquad 1 \le i \le n,$$

(ii)

$$\sum_{\substack{i=1 \\ i\neq j}}^{n} \frac{|a_{i,j}|}{|a_{i,i}|} < 1, \qquad 1 \le j \le n,$$

(iii) stĺpcové kritérium

$$\sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{i,j}| < |a_{j,j}|, \qquad 1 \le j \le n,$$

tak platí tvrdenie vety 7.4.

DÔKAZ: Najprv dokážme (i). Napíšme si prvkové vyjadrenie pre maticu $\boldsymbol{C}$.

$$\boldsymbol{C} = -\begin{pmatrix} 0 & \frac{a_{1,2}}{a_{1,1}} & \frac{a_{1,3}}{a_{1,1}} & \dots & \frac{a_{1,n}}{a_{1,1}} \\ \frac{a_{2,1}}{a_{2,2}} & 0 & \frac{a_{2,3}}{a_{2,2}} & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{a_{n-1,1}}{a_{n-1,n-1}} & \dots & \frac{a_{n-1,n-2}}{a_{n-1,n-1}} & 0 & \frac{a_{n-1,n}}{a_{n-1,n-1}} \\ \frac{a_{n,1}}{a_{n,n}} & \dots & \dots & \frac{a_{n,n-1}}{a_{n,n}} & 0 \end{pmatrix}.$$

Takže

$$\max_{1\le i\le n} \sum_{j=1}^{n} |c_{i,j}| = \max_{1\le i\le n} \sum_{\substack{j=1\\ j\ne i}}^{n} \left|\frac{a_{i,j}}{a_{i,i}}\right|$$

predstavuje normu matice $\boldsymbol{C}$. Ak je táto menšia ako 1, potom platí tvrdenie vety 7.4. Poznamenajme, že (i) znamená riadkovú diagonálnu dominantnosť matice $\boldsymbol{A}$.

V časti (ii) je použitá len iná norma pre maticu $\boldsymbol{C}$.

Dôkaz časti (iii) je trochu komplikovanejší. Najprv si pripomeňme isté fakty z teórie matíc. Ak $\boldsymbol{M}$ je štvorcová matica, $\boldsymbol{M}^T$ transponovaná matica, tak

(a) $\|\boldsymbol{M}\| = \|\boldsymbol{M}^T\|$.

(b) Ak $\boldsymbol{D}$ je diagonálna matica, tak

$$\boldsymbol{D}\boldsymbol{M}\boldsymbol{D}^{-1}\boldsymbol{x} = \lambda\boldsymbol{x} \Longleftrightarrow \boldsymbol{M}\boldsymbol{D}^{-1}\boldsymbol{x} = \boldsymbol{D}^{-1}\lambda\boldsymbol{x} = \lambda\boldsymbol{D}^{-1}\boldsymbol{x},$$

t.j. vlastné čísla matíc $\boldsymbol{D}\boldsymbol{M}\boldsymbol{D}^{-1}$ a $\boldsymbol{M}$ sú rovnaké.

Vzťah (iii) vyjadruje, že diagonála je dominantná v každom stĺpci matice $\boldsymbol{A}$, t.j. diagonála je dominantná v každom riadku matice $\boldsymbol{A}^T$. Použime teraz už dokázanú časť (i). Urobme $\boldsymbol{LDU}$-rozklad matice $\boldsymbol{A}^T$. Platí

$$\boldsymbol{A}^T = -\boldsymbol{L}_1 + \boldsymbol{D}_1 - \boldsymbol{U}_1, \text{pričom} \quad \boldsymbol{L}_1 = \boldsymbol{U}^T, \quad \boldsymbol{D}_1 = \boldsymbol{D}, \quad \boldsymbol{U}_1 = \boldsymbol{L}^T.$$

Ak platí (iii), potom z (i) vieme, že $\|\boldsymbol{D}^{-1}(\boldsymbol{L}_1 + \boldsymbol{U}_1)\| < 1$, pričom sme použili riadkovú normu $\|\cdot\|$. Ďalej

$$\boldsymbol{D}^{-1}(\boldsymbol{L}_1 + \boldsymbol{U}_1) = \boldsymbol{D}^{-1}(\boldsymbol{U}^T + \boldsymbol{L}^T) = \boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})^T = [(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{D}^{-1}]^T.$$

Teda $1 > \|[(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{D}^{-1}]^T\| = \|(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{D}^{-1}\| = \|\boldsymbol{D}\boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{D}^{-1}\|$. Odtiaľ vyplýva, že všetky vlastné čísla matice $\boldsymbol{D}\boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{D}^{-1}$ sú v absolútnej hodnote menšie ako 1, čiže i vlastné čísla matice $\boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})$ sú v absolútnej hodnote menšie ako 1. Použitím vety 7.1 dostaneme požadované tvrdenie. □

7.3 Gaussova-Seidelova iteračná metóda

Nech $\boldsymbol{A}$ je regulárna matica rádu $n \times n$, pričom jej diagonálne prvky sú nenulové. Urobme $\boldsymbol{LDU}$-rozklad matice $\boldsymbol{A}$ tak, ako aj pri Jacobiho iteračnej metóde, t.j. $\boldsymbol{A} = -\boldsymbol{L} + \boldsymbol{D} - \boldsymbol{U}$. Vyjdime opäť zo vzťahu (7.1). Tento si teraz modifikujme takto

$$\boldsymbol{x} = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}\boldsymbol{x} + (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{b}.$$

Zaveďme iterácie nasledujúcim spôsobom

$$\boldsymbol{x}_{k+1} = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}\boldsymbol{x}_k + (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{b}. \tag{7.6}$$

Ľahko vidieť, že sa jedná o modifikáciu metódy postupných aproximácií, pričom $\boldsymbol{C} = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}$ a $\boldsymbol{g} = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{b}$. Takže ak bude $\|\boldsymbol{C}\| = \|(\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}\| < 1$, tak bude metóda konvergovať pre ľubovoľnú štartovaciu hodnotu $\boldsymbol{x}_0$.

Odvoďme si teraz formulku pre výpočet súradníc vektoru $\boldsymbol{x}_{k+1}$. Zo vzťahu (7.6) vyplýva

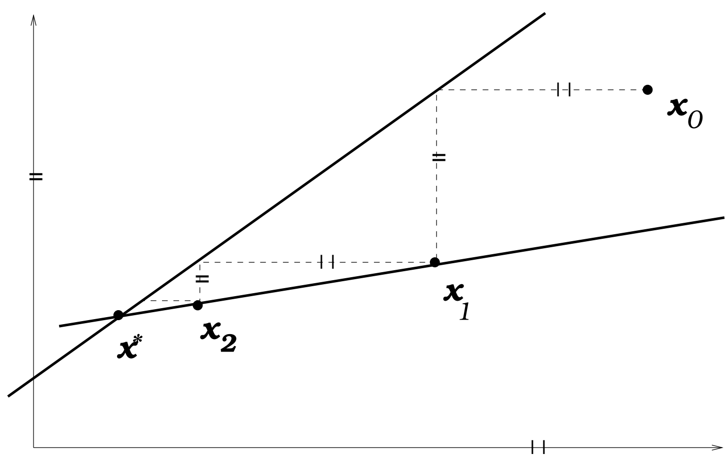
$$\begin{aligned}
(\boldsymbol{D} - \boldsymbol{L})\boldsymbol{x}_{k+1} &= \boldsymbol{U}\boldsymbol{x}_k + \boldsymbol{b} \\
\boldsymbol{D}\boldsymbol{x}_{k+1} &= \boldsymbol{L}\boldsymbol{x}_{k+1} + \boldsymbol{U}\boldsymbol{x}_k + \boldsymbol{b} \\
\boldsymbol{x}_{k+1} &= \boldsymbol{D}^{-1}\left[\boldsymbol{L}\boldsymbol{x}_{k+1} + \boldsymbol{U}\boldsymbol{x}_k + \boldsymbol{b}\right].
\end{aligned}$$

Ak označíme súradnice n-rozmerného vektora ako $\boldsymbol{y} = (y^1, \ldots, y^n)^T$, tak môžeme písať

$$\boldsymbol{x}_{k+1}^i = -\frac{1}{a_{i,i}}\left[\sum_{j=1}^{i-1} a_{i,j}x_{k+1}^j + \sum_{j=i+1}^{n} a_{i,j}x_k^j - b_i\right]. \tag{7.7}$$

Gaussova-Seidelova metóda sa niekedy nazýva i *jednokrokovým cyklickým procesom.* Pri výpočte i-tej súradnice $(k+1)$-vej aproximácie sa do pravej strany dosadzujú už vypočítané súradnice $x_{k+1}^1, \ldots, x_{k+1}^{i-1}$ tejto $(k+1)$-vej aproximácie.

Gaussova-Seidelova iteračná metóda pre sústavu dvoch rovníc je graficky znázornená na obrázku 7.2.



Obr. 7.2. Gaussova-Seidelova metóda v $\mathbb{R}^2$

Príklad 7.1 Dané sú matice

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 1 & 1 & 1 \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}.$$

Preverte konvergenciu Gaussovej-Seidelovej i Jacobiho metódy pre obe matice.

Riešenie: Uvažujme najprv maticu $\boldsymbol{A}$. Zrejme

$$\boldsymbol{A} = \boldsymbol{D}_A - \boldsymbol{L}_A - \boldsymbol{U}_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -2 & -2 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Iteračná matica pre Jacobiho metódu je $\boldsymbol{D}_A^{-1}(\boldsymbol{L}_A + \boldsymbol{U}_A)$. Pre charakteristický polynóm platí

$$p_J(\lambda) = \det\left[\boldsymbol{D}_A^{-1}(\boldsymbol{L}_A + \boldsymbol{U}_A) - \lambda \boldsymbol{I}\right] = \begin{vmatrix} -\lambda & -2 & 2 \\ -1 & -\lambda & -1 \\ -2 & -2 & -\lambda \end{vmatrix} = -\lambda^3.$$

Rovnica $p_J(\lambda) = 0$ má jediné riešenie $\lambda = 0$ a spektrálny rádius iteračnej matice pre Jacobiho metódu je 0, teda metóda konverguje pre ľubovoľný štartovací vektor.

Iteračná matica pre Gaussovu-Seidelovu metódu je $(\boldsymbol{D}_A - \boldsymbol{L}_A)^{-1}\boldsymbol{U}_A$. Pre charakteristický polynóm platí

$$p_G(\lambda) = \det\left[(\boldsymbol{D}_A - \boldsymbol{L}_A)^{-1}\boldsymbol{U}_A - \lambda \boldsymbol{I}\right] = \begin{vmatrix} -\lambda & -2 & 2 \\ 0 & 2-\lambda & -3 \\ 0 & 0 & 2-\lambda \end{vmatrix} = -\lambda(2-\lambda)^2.$$

Riešenia rovnice $p_G(\lambda) = 0$ sú $\lambda = 0$ a $\lambda = 2$. Spektrálny rádius iteračnej matice pre Gaussovu-Seidelovu metódu je 2, takže metóda nekonverguje.

Teraz uvažujme maticu $\boldsymbol{B}$. Platí

$$\boldsymbol{B} = \boldsymbol{D}_B - \boldsymbol{L}_B - \boldsymbol{U}_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} - \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Pre charakteristický polynóm Jacobiho metódy dostaneme

$$p_J(\lambda) = \det\left[\boldsymbol{D}_B^{-1}(\boldsymbol{L}_B + \boldsymbol{U}_B) - \lambda \boldsymbol{I}\right] = \begin{vmatrix} -\lambda & \frac{1}{2} & \frac{1}{2} \\ -1 & -\lambda & -1 \\ \frac{1}{2} & \frac{1}{2} & -\lambda \end{vmatrix} = -\left(\lambda^2 - \frac{\lambda}{2} + 1\right)\left(\frac{1}{2} + \lambda\right).$$

Rovnica $p_J(\lambda) = 0$ má tri riešenia $\lambda = -\frac{1}{2}$, $\lambda = \frac{1}{4} \pm \frac{i\sqrt{15}}{4}$, takže spektrálny rádius iteračnej matice sa rovná 1. Jacobiho metóda nekonverguje.

Pre charakteristický polynóm Gaussovej-Seidelovej metódy platí

$$p_G(\lambda) = \det\left[(\boldsymbol{D}_B - \boldsymbol{L}_B)^{-1}\boldsymbol{U}_B - \lambda \boldsymbol{I}\right] = \begin{vmatrix} -\lambda & \frac{1}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} - \lambda & -\frac{3}{2} \\ 0 & 0 & -\frac{1}{2} - \lambda \end{vmatrix} = -\frac{\lambda}{4}(1 + 2\lambda)^2.$$

Riešenia rovnice $p_G(\lambda) = 0$ sú $\lambda = 0$ a $\lambda = -\frac{1}{2}$. Spektrálny rádius iteračnej matice pre Gaussovu-Seidelovu metódu je $\frac{1}{2}$, takže metóda konverguje pre ľubovoľný štartovací vektor. □

7.4 Metóda konjugovaných gradientov

Chceme riešiť systém lineárnych algebraických rovníc (7.1) v $\mathbb{R}^n$ pričom matica systému $\boldsymbol{A}$ je symetrická a pozitívne definitná. Základom pre odvodenie algoritmu konjugovaných gradientov[1] je nasledujúca veta.

[1] V literatúre sa často označuje ako *CG - metóda.* Skratka pochádza z angličtiny "conjugate gradients method".

Veta 7.5 *Vektor $\boldsymbol{x}$ je riešením systému (7.1) so symetrickou pozitívne definitnou maticou $\boldsymbol{A}$ vtedy a len vtedy, ak je minimom energetického funkcionálu*

$$F(\boldsymbol{v}) = \frac{1}{2}(\boldsymbol{v}^T, \boldsymbol{A}\boldsymbol{v}) - (\boldsymbol{b}^T, \boldsymbol{v}).$$

DÔKAZ: Jednoduchým výpočtom sa presvedčíme, že pre gradient funkcie F platí

$$\nabla F(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v} - \boldsymbol{b},$$

kde gradient je vektor prvých parciálnych derivácii. Riešenie $\boldsymbol{x}$ systému (7.1) spĺňa nutnú podmienku pre existenciu extrému. Hessova matica pre $F(\boldsymbol{v})$, t.j. matica druhých parciálnych derivácií, sa rovná matici $\boldsymbol{A}$ a tá je pozitívne definitná. Teda $\boldsymbol{x}$ je minimom $F(\boldsymbol{v})$. Naopak každé minimum $\boldsymbol{v}_0$ kvadratického funkcionálu $F(\boldsymbol{v})$ je riešením (7.1). Z jednoznačnosti riešenia (7.1) vyplýva $\boldsymbol{v}_0 = \boldsymbol{x}$. □

Z vety 7.5 je zrejmé, že riešiť (7.1) znamená nájsť minimum energetického funkcionálu. Opíšme teraz iteračný spôsob hľadania minima $F(\boldsymbol{v})$. Vyjdime pritom zo štartovacieho vektora $\boldsymbol{x}_0$ a hľadajme minimum $F(\boldsymbol{v})$ v danom smere $\boldsymbol{p} \neq \boldsymbol{0}$. Platí

$$\begin{aligned} F(\boldsymbol{x}_0 + t\boldsymbol{p}) &= \frac{1}{2}(\boldsymbol{x}_0^T + t\boldsymbol{p}^T, \boldsymbol{A}(\boldsymbol{x}_0 + t\boldsymbol{p})) - (\boldsymbol{b}^T, \boldsymbol{x}_0 + t\boldsymbol{p}) \\ &= \frac{1}{2}t^2(\boldsymbol{p}^T, \boldsymbol{A}\boldsymbol{p}) + t(\boldsymbol{p}^T, \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}) + F(\boldsymbol{x}_0). \end{aligned}$$

Zo vzťahu $\dfrac{\partial F(\boldsymbol{x}_0 + t\boldsymbol{p})}{\partial t} = 0$ vidíme, že minimum sa nadobúda pre

$$t_{min} = -\frac{(\boldsymbol{p}^T, \boldsymbol{r})}{(\boldsymbol{p}^T, \boldsymbol{A}\boldsymbol{p})}, \tag{7.8}$$

pričom $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$ je *reziduálny vektor*. Skutočne pre hodnotu t_{min} dostaneme minimum $F(\boldsymbol{v})$ v smere vektora $\boldsymbol{p}$ ak vychádzame z bodu $\boldsymbol{x}_0$. Tu je však potrebné aby $(\boldsymbol{p}^T, \boldsymbol{r}) \neq 0$, lebo inak by sme sa nepohli z miesta $\boldsymbol{x}_0$. Nová aproximácia riešenia bude $\boldsymbol{x}_1 = \boldsymbol{x}_0 + t_{min}\boldsymbol{p}$.

Lema 7.1 *Nech $\boldsymbol{r}_1 = \boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}$ je reziduálny vektor v $\boldsymbol{x}_1 = \boldsymbol{x}_0 + t_{min}\boldsymbol{p}$. Potom*

$$(\boldsymbol{p}^T, \boldsymbol{r}_1) = 0.$$

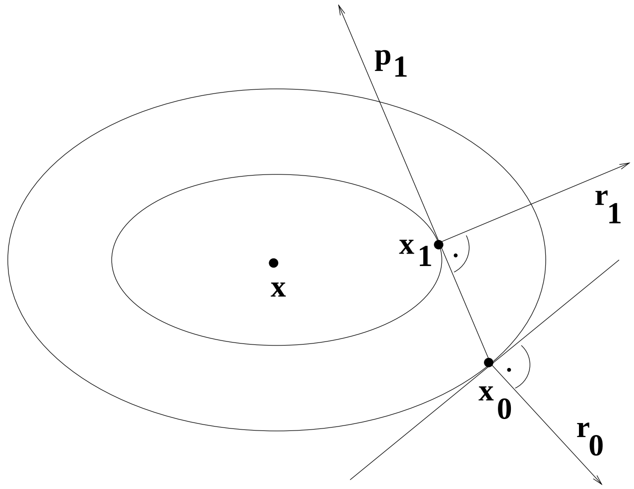
DÔKAZ: Pomocou (7.8) dostaneme

$$(\boldsymbol{p}^T, \boldsymbol{r}_1) = (\boldsymbol{p}^T, \boldsymbol{A}(\boldsymbol{x}_0 + t_{min}\boldsymbol{p}) - \boldsymbol{b}) = (\boldsymbol{p}^T, \boldsymbol{r}) + t_{min}(\boldsymbol{p}^T, \boldsymbol{A}\boldsymbol{p}) = 0.$$

Teda reziduálny vektor v bode $\boldsymbol{x}_1$ je ortogonálny k pôvodnému smeru $\boldsymbol{p}$. □

Tvrdenie tejto lemy sa dá geometricky pre $n = 2$ interpretovať takto. Potenciálové krivky $F(\boldsymbol{v}) = \textit{konštanta}$ sú znázornené elipsami so spoločným priesečníkom ich osí bodom $\boldsymbol{x}$, ktorý predstavuje exaktné riešenie sústavy rovníc. Začneme v bode $\boldsymbol{x}_0$ a zvolíme smer $\boldsymbol{p}_1$. V bode $\boldsymbol{x}_0$ je reziduálny vektor kolmý na potenciálovú krivku prechádzajúcu cez bod $\boldsymbol{x}_0$. Nasledujúcu aproximáciu $\boldsymbol{x}_1$ dostaneme ako minimum $F(\boldsymbol{v})$ v danom smere $\boldsymbol{p}_1$. Podľa lemy je $\boldsymbol{r}_1$ kolmý na $\boldsymbol{p}_1$, teda $\boldsymbol{p}_1$ je vlastne dotyčnicou ku potenciálovej krivke prechádzajúcej cez $\boldsymbol{x}_1$ - pozri obrázok 7.3.

Zaveďme nasledujúcu terminológiu.



Obr. 7.3. Geometrická interpretácia lemy 7.1

Definícia 7.4 *Dva vektory* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ *sa nazývajú* konjugované *alebo* $\boldsymbol{A}$-ortogonálne, *ak pre pozitívne definitnú maticu* $\boldsymbol{A}$ *platí*

$$(\boldsymbol{u}^T, \boldsymbol{A}\boldsymbol{v}) = 0.$$

ALGORITMUS

Vstup štartovací vektor $\boldsymbol{x}_0$
rezíduum $\boldsymbol{r}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$
prvý smer $\boldsymbol{p}_1 = -\boldsymbol{r}_0$
pre $k = 1, 2, \dots$

$$\begin{aligned}
\boldsymbol{x}_k &= \boldsymbol{x}_{k-1} + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}\, \boldsymbol{p}_k, \\
\boldsymbol{r}_k &= \boldsymbol{r}_{k-1} + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}\, \boldsymbol{A}\boldsymbol{p}_k, \\
\boldsymbol{p}_{k+1} &= -\boldsymbol{r}_k + \frac{(\boldsymbol{r}_k^T, \boldsymbol{r}_k)}{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}\, \boldsymbol{p}_k.
\end{aligned} \tag{7.9}$$

Vlastnosti CG-metódy

Opíšme si najdôležitejšie vlastnosti metódy konjugovaných gradientov. Tieto budeme potrebovať pri dôkaze konvergencie algoritmu. Tvrdenie lemy 7.1 o ortogonalite reziduálneho a smerového vektoru sa dá zovšeobecniť i pre ďalšie kroky.

Veta 7.6 *Reziduálne vektory* $\boldsymbol{r}_k$ *tvoria ortogonálny systém. Smerové vektory* $\boldsymbol{p}_k$ *sú po dvojiciach konjugované. Teda pre* $k \geq 2$ *platí*

$$\begin{aligned}
(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_j) &= 0, && \textit{pre } j = 0, \dots, k-2; \\
(\boldsymbol{r}_{k-1}^T, \boldsymbol{p}_j) &= 0, && \textit{pre } j = 1, \dots, k-1; \\
(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_j) &= 0, && \textit{pre } j = 1, \dots, k-1.
\end{aligned} \tag{7.10}$$

DÔKAZ: Použijeme matematickú indukciu podľa k. Začnime s $k = 2$. Potom pomocou lemy 7.1 máme

$$(\boldsymbol{r}_1^T, \boldsymbol{r}_0) = -(\boldsymbol{r}_1^T, \boldsymbol{p}_1) = 0.$$

Z konštrukcie algoritmu vyplýva

$$\begin{aligned}
(\boldsymbol{p}_2^T, \boldsymbol{A}\boldsymbol{p}_1) &= \left(-\boldsymbol{r}_1^T + \frac{(\boldsymbol{r}_1^T, \boldsymbol{r}_1)}{(\boldsymbol{r}_0^T, \boldsymbol{r}_0)}\boldsymbol{p}_1^T, \boldsymbol{A}\boldsymbol{p}_1\right) \\
&= -(\boldsymbol{r}_1^T, \boldsymbol{A}\boldsymbol{p}_1) + \frac{(\boldsymbol{r}_1^T, \boldsymbol{r}_1)}{(\boldsymbol{r}_0^T, \boldsymbol{r}_0)}(\boldsymbol{p}_1^T, \boldsymbol{A}\boldsymbol{p}_1) \\
&= -\left(\boldsymbol{r}_1^T, [\boldsymbol{r}_1 - \boldsymbol{r}_0]\frac{(\boldsymbol{p}_1^T, \boldsymbol{A}\boldsymbol{p}_1)}{(\boldsymbol{r}_0^T, \boldsymbol{r}_0)}\right) + \frac{(\boldsymbol{r}_1^T, \boldsymbol{r}_1)}{(\boldsymbol{r}_0^T, \boldsymbol{r}_0)}(\boldsymbol{p}_1^T, \boldsymbol{A}\boldsymbol{p}_1) \\
&= 0,
\end{aligned}$$

lebo $(\boldsymbol{r}_1^T, \boldsymbol{r}_0) = 0$. Nech sú splnené vzťahy (7.10) pre k. Dokážme ich platnosť pre $k+1$. Pre reziduálne vektory $(j \geq 1)$ platí

$$\begin{aligned}
(\boldsymbol{r}_j^T, \boldsymbol{r}_k) &= \left(\boldsymbol{r}_j^T, \boldsymbol{r}_{k-1} + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}\boldsymbol{A}\boldsymbol{p}_k\right) \\
&= (\boldsymbol{r}_j^T, \boldsymbol{r}_{k-1}) + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}(\boldsymbol{r}_j^T, \boldsymbol{A}\boldsymbol{p}_k) \\
&= (\boldsymbol{r}_j^T, \boldsymbol{r}_{k-1}) + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}\left(\frac{(\boldsymbol{r}_j^T, \boldsymbol{r}_j)}{(\boldsymbol{r}_{j-1}^T, \boldsymbol{r}_{j-1})}\boldsymbol{p}_j^T - \boldsymbol{p}_{j+1}^T, \boldsymbol{A}\boldsymbol{p}_k\right) \\
&= (\boldsymbol{r}_j^T, \boldsymbol{r}_{k-1}) + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})(\boldsymbol{r}_j^T, \boldsymbol{r}_j)}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)(\boldsymbol{r}_{j-1}^T, \boldsymbol{r}_{j-1})}(\boldsymbol{p}_j^T, \boldsymbol{A}\boldsymbol{p}_k) \\
&\quad - \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}(\boldsymbol{p}_{j+1}^T, \boldsymbol{A}\boldsymbol{p}_k).
\end{aligned}$$

Ak $j = 1, \ldots, k-2$, potom sa všetky tri posledné sčítance podľa indukčného predpokladu rovnajú nule. Ak $j = k-1$, tak zo symetrie matice $\boldsymbol{A}$ vyplýva $(\boldsymbol{p}_j^T, \boldsymbol{A}\boldsymbol{p}_k) = (\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_j) = 0$ a súčet prvého a tretieho sčítanca je nula. Pre $j = 0$ postupujeme podobne a dostaneme

$$(\boldsymbol{r}_0^T, \boldsymbol{r}_k) = (\boldsymbol{r}_0^T, \boldsymbol{r}_{k-1}) + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}(\boldsymbol{r}_0^T, \boldsymbol{A}\boldsymbol{p}_k) = 0,$$

lebo $\boldsymbol{r}_0 = -\boldsymbol{p}_1$ v druhom sčítanci. Teda spolu zhrnuté

$$(\boldsymbol{r}_k^T, \boldsymbol{r}_j) = 0, \qquad \text{pre } j = 0, \ldots, k-1. \tag{7.11}$$

Ďalej

$$(\boldsymbol{p}_j^T, \boldsymbol{r}_k) = (\boldsymbol{p}_j^T, \boldsymbol{r}_{k-1}) + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_k)}\ (\boldsymbol{p}_j^T, \boldsymbol{A}\boldsymbol{p}_k).$$

Pre $j = 1, \ldots, k-1$ pomocou indukčného predpokladu dostaneme 0. Ak $j = k$ potom pomocou rekurzie dostaneme

$$(\boldsymbol{r}_k^T, \boldsymbol{p}_k) = -(\boldsymbol{r}_k^T, \boldsymbol{r}_{k-1}) + \frac{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}{(\boldsymbol{r}_{k-2}^T, \boldsymbol{r}_{k-2})}(\boldsymbol{r}_k^T, \boldsymbol{p}_{k-1}) = 0.$$

Pre smery máme

$$\begin{aligned}
(\boldsymbol{p}_{k+1}^T, \boldsymbol{A}\boldsymbol{p}_j) &= -(\boldsymbol{r}_k^T, \boldsymbol{A}\boldsymbol{p}_j) + \frac{(\boldsymbol{r}_k^T, \boldsymbol{r}_k)}{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_j) \\
&= -\left(\boldsymbol{r}_k^T, [\boldsymbol{r}_j - \boldsymbol{r}_{j-1}]\frac{(\boldsymbol{p}_j^T, \boldsymbol{A}\boldsymbol{p}_j)}{(\boldsymbol{r}_{j-1}^T, \boldsymbol{r}_{j-1})}\right) + \frac{(\boldsymbol{r}_k^T, \boldsymbol{r}_k)}{(\boldsymbol{r}_{k-1}^T, \boldsymbol{r}_{k-1})}(\boldsymbol{p}_k^T, \boldsymbol{A}\boldsymbol{p}_j) \\
&= 0.
\end{aligned}$$

□

O konvergencii metódy hovorí nasledujúca veta.

Veta 7.7 *Metóda konjugovaných gradientov konverguje ku riešeniu sústavy lineárnych algebraických rovníc o n neznámych najneskôr po n krokoch.*

DÔKAZ: Reziduálne vektory $\boldsymbol{r}_0, \ldots, \boldsymbol{r}_k$ tvoria v $\mathbb{R}^n$ ortogonálny systém. Teda najneskôr po n krokoch generujú celý $\mathbb{R}^n$. Takže $\boldsymbol{r}_n = 0$, čo znamená, že $\boldsymbol{x}_n$ je riešením (7.1). □

Teoreticky je takto skoncipovaný CG-algoritmus vždy konečný proces. Dokonca tu nezáleží ani na voľbe štartovacieho vektora $\boldsymbol{x}_0$. V skutočnosti tomu tak nie je. Z hľadiska numerických chýb nie je možné splniť podmienky ortogonality reziduálnych vektorov presne. Takže vykonávanie algoritmu i pre $k > n$ môže byť zmysluplné. V praxi však majú sústavy lineárnych algebraických rovníc obrovské rozmery (n môže byť viac ako 10^6), takže tu ostáva dúfať, že CG-metóda bude aj pre malý počet iterácií dávať dobrú aproximáciu exaktného riešenia. V tomto prípade voľba $\boldsymbol{x}_0$ môže zohrať dôležitú úlohu.

Cvičenie 7.1 Nech

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

Pre ktoré hodnoty $\boldsymbol{x} \in \mathbb{R}^3$ má nekonečný rad $\sum_{i=0}^{\infty} \frac{1}{i!}\boldsymbol{A}^i\boldsymbol{x}$ konečnú normu? Uvažujte rôzne typy vektorových noriem.

Cvičenie 7.2 Nech $n \in \mathbb{N}, \alpha \in \mathbb{R}$. Uvažujme maticu otočenia o uhol α v rovine danú

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}.$$

Pre dané $\boldsymbol{b} \in \mathbb{R}^2$ chceme riešiť systém

$$A^n\boldsymbol{x} = \boldsymbol{b}.$$

Pre aké n bude Gaussova-Seidelova iteračná metóda konvergovať nezávisle od voľby štartovacieho vektora $\boldsymbol{x}_0$?

Cvičenie 7.3 Charakterizujte všetky regulárne matice $A \in \mathbb{R}^{n\times n}$, pre ktoré Jacobiho iteračná metóda už v prvom kroku dáva presné riešenie, ak na začiatku štartovací vektor bol $\boldsymbol{x}_0 = \boldsymbol{0}$.

Kapitola 8

Numerická integrácia

Viaceré problémy aplikovanej matematiky vedú popri inom i k výpočtom integrálov, ktoré sa nedajú vyjadriť v explicitnej forme. Tieto sa počítajú približnými metódami, čiže numerickou integráciou resp. kvadratúrou. S početnými aplikáciami kvadratúry sa stretávame pri výpočte plôch, objemov, v teórii pravdepodobnosti, pri riešení integrálnych i diferenciálnych rovníc atď. Na výpočet určitého integrálu sa dajú použiť viaceré známe spôsoby. Výber vhodnej metódy záleží od rôznych kritérií:

- Hladkosť integrovanej funkcie.
- Dostupné informácie o integrovanej funkcii, napr. či je daná v tabelovanej forme alebo v spojitom tvare.
- Požadovaná presnosť numerickej integrácie. [1]
- Koľkokrát sa kvadratúra má vykonať. [2]

V tejto kapitole si ukážeme niektoré spôsoby numerického výpočtu integrálov, ktoré sú založené na interpolácii integrovanej funkcie polynómom.

8.1 Interpolačné kvadratúrne metódy

V kapitole 4 sme sa zaoberali aproximáciou funkcie pomocou Lagrangeových polynómov. Tieto mali v bodoch $x_0, \ldots, x_n$ tie isté hodnoty ako daná funkcia f. Uvažujme teraz prípad, že $x_0, \ldots, x_n$ sú rovnomerne rozložené, usporiadané vzostupne s krokom h. Teda

$$x_k = x_0 + kh, \qquad k = 0, 1, \ldots, n.$$

Zavedením substitúcie

$$x = x_0 + th$$

dostaneme

[1] Tu si treba uvedomiť, že samotná kvadratúra často býva len dielčím problémom. Preto jej presnosť musí byť vybalancovaná s presnosťou iných numerických schém použitých pri riešení problému.

[2] Pri numerickom riešení parciálnych diferenciálnych rovníc napr. metódou konečných prvkov je pre každý element (a tých môžu byť tisíce) potrebné viac razy previesť kvadratúru. Takže v tomto prípade treba upustiť od prehnanej presnosti danej numerickej integrácie, lebo inak celkový čas výpočtu riešenia problému bude neúnosne dlhý.

$$\frac{x - x_i}{x_j - x_i} = \frac{t - i}{j - i} \quad .$$

Takže pre Lagrangeove koeficienty môžeme písať

$$l_j(x) = \prod_{\substack{i=0 \\ i \neq j}}^{n} \frac{x - x_i}{x_j - x_i} = \prod_{\substack{i=0 \\ i \neq j}}^{n} \frac{t - i}{j - i} = \frac{(-1)^{n-j}}{j!(n-j)!} \prod_{\substack{i=0 \\ i \neq j}}^{n} (t - i).$$

Lagrangeov tvar interpolačného polynómu bude mať tvar (pozri (4.4))

$$p(t) = \sum_{j=0}^{n} l_j(t) f_j = \sum_{j=0}^{n} f(x_j) \frac{(-1)^{n-j}}{j!(n-j)!} \prod_{\substack{i=0 \\ i \neq j}}^{n} (t - i). \tag{8.1}$$

Pre chybu Lagrangeovej interpolácie z vety 4.2 vyplýva

$$f(x_0 + th) - p(t) = t(t-1)\dots(t-n)h^{n+1} \frac{f^{(n+1)}(\xi)}{(n+1)!} \tag{8.2}$$

pre nejaké $x_0 \le \xi \le x_n$.

Príklad 8.1 Ak $n = 1$, tak $p(t) = (1-t)f_0 + tf_1$. Takže dostaneme lineárnu interpoláciu funkcie f. Pre jej chybu máme

$$f(x_0 + th) - p(t) = h^2 f^{(2)}(\xi) \frac{t(t-1)}{2}$$

pre $x_0 \le \xi \le x_1$. □

Interpolačné kvadratúrne metódy sú založené na aproximácii funkcie f jej interpolačným polynómom p. Teda

$$\int_a^b f(x)dx \approx \int_a^b p(x)dx.$$

Pre jednoduchosť uvažujme len prípad, že krajné body a, b splývajú s niektorými z bodov $x_0, \dots, x_n$, napr. $a = x_p$, $b = x_q$ pre $p < q$. Integrovaním rovnosti $f(x) = p(x) + E(x)$, kde $E(x)$ je chyba interpolácie, dostaneme

$$\int_{x_p}^{x_q} f(x)dx = \int_{x_p}^{x_q} p(x)dx + \int_{x_p}^{x_q} E(x)dx = \int_{x_p}^{x_q} p(x)dx + e,$$

kde e je chybou numerickej integrácie. Teda

$$\int_{x_p}^{x_q} f(x)dx = \sum_{j=0}^{n} A_j f(x_j) h \ + \ e, \tag{8.3}$$

pričom

$$A_j = \frac{(-1)^{n-j}}{j!(n-j)!} \int_p^q \prod_{\substack{i=0 \\ i \neq j}}^{n} (t-i)dt. \tag{8.4}$$

Koeficienty A_j môžu byť vo všeobecnosti i záporné. Vzorce typu (8.3) sa nazývajú *Newtonove-Cotesove.* Tieto môžu byť otvoreného alebo uzavretého typu podľa toho, či krajné body intervalu sú považované za uzly alebo nie.

8.1.1 Newtonove-Cotesove vzorce uzavretého typu

Nech $p = 0$, $q = n$. To znamená, že integračný aj interpolačný interval sú zhodné, t.j. koncové body intervalu berieme za uzly. Podrobnejšou analýzou chyby sa dá dokázať nasledujúce tvrdenie.

Veta 8.1 *Nech funkcia f má $(n+2)$ spojitých derivácií v intervale $[a,b]$. Potom pre chybu Newtonovej-Cotesovej metódy s rovnomerne rozloženými uzlami $x_k = a+kh$, $k = 0,\dots,n$, $h = \dfrac{b-a}{n}$ platí*

- *ak n je párne*

$$e = \int_a^b f(x)dx - \sum_{j=0}^{n} A_j f(x_j)h = h^{n+3}\frac{f^{(n+2)}(\xi)}{(n+2)!}\int_0^n t\prod_{i=0}^{n}(t-i)dt,$$

- *ak n je nepárne*

$$e = \int_a^b f(x)dx - \sum_{j=0}^{n} A_j f(x_j)h = h^{n+2}\frac{f^{(n+1)}(\xi)}{(n+1)!}\int_0^n \prod_{i=0}^{n}(t-i)dt,$$

pričom $a < \xi < b$.

DÔKAZ: Podrobný dôkaz je v [8], [3]. □

To znamená, že uzavretý Newtonov-Cotesov vzorec je pre nepárne n presný pre polynómy stupňa n, pre párne n je kvadratúrny vzorec presný pre polynómy stupňa $(n+1)$.

Lichobežníková metóda

Newtonova-Cotesova metóda uzavretého typu pre $n = 1$ sa nazýva *lichobežníková metóda.* Ide teda o metódu, ktorej základom je lineárna interpolácia.

$$\begin{aligned}\int_a^b f(x)dx &\approx \int_a^b \left[f(a) + \frac{x-a}{b-a}(f(b)-f(a))\right]dx \\ &= \frac{f(a)+f(b)}{2}(b-a).\end{aligned}$$

Pre chybu kvadratúry dostaneme pre $h = b - a$

$$e = -\frac{h^3}{12}f^{(2)}(\xi), \qquad \xi \in (a,b).$$

Simpsonova metóda

Newtonova-Cotesova metóda uzavretého typu pre $n = 2$ sa nazýva *Simpsonova metóda.* Jej základom je kvadratická interpolácia. Uzly sa volia takto

$$x_0 = a, \quad x_1 = \frac{a+b}{2}, \quad x_2 = b,$$

pričom $h = \frac{b-a}{2}$.

Pre približný výpočet integrálu dostaneme

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right].$$

Pre chybu kvadratúry máme

$$e = -\frac{h^5}{90} f^{(4)}(\xi), \qquad \xi \in (a,b).$$

Napriek tomu, že sme pri interpolácii vyšli z kvadratickej interpolácie, ľahko sa dá overiť presnosť Simpsonovej metódy i pre polynómy tretieho stupňa. Stačí to preveriť pre polynóm x^3.

$$\frac{b-a}{6}\left[a^3 + 4\left(\frac{a+b}{2}\right)^3 + b^3\right] = \frac{b^4 - a^4}{4} = \int_a^b x^3 dx.$$

8.1.2 Newtonove-Cotesove vzorce otvoreného typu

Pri metóde otvoreného typu sa krajné body intervalu, cez ktorý integrujeme, neberú ako uzly. V tomto prípade budú uzly symetricky rozložené podľa stredu intervalu. Teda stačí položiť $p = -1$, $q = n + 1$ v (8.3), (8.4). Pre analýzu kvadratúrnej chyby platí analogická veta k vete 8.1.

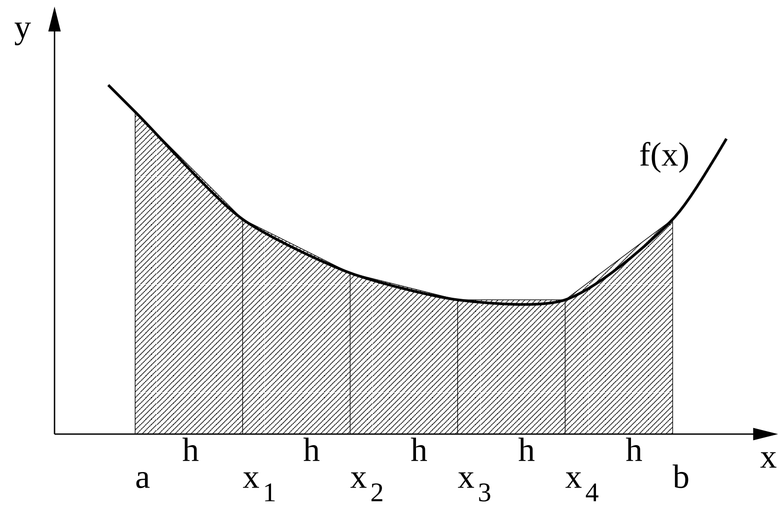
Obdĺžniková metóda

Špeciálny prípad ak $n = 0$ sa nazýva *obdĺžniková metóda.* Pre ňu platí

$$\int_a^b f(x)dx = (b-a)\ f\left(\frac{a+b}{2}\right) + \frac{h^3}{3} f^{(2)}(\xi),$$

pre $\xi \in (a,b)$, $h = \frac{b-a}{2}$. Pri praktickom výpočte hodnoty určitého integrálu sa dĺžka intervalu vynásobí hodnotou funkcie v strede intervalu.

Poznámka 8.1 *Všimnime si, že ku každému kvadratúrnemu vzorcu s uzlami x_k sa dá nájsť spojitá funkcia na danom intervale taká, že približná hodnota určitého integrálu bude rovná 0. Na to stačí, aby sa hodnoty funkcie v uzloch x_k rovnali 0.* □



Obr. 8.1. Lichobežníková metóda

8.1.3 Zložené kvadratúrne metódy

Ak chceme interpolačnými metódami dosiahnuť požadovanú presnosť, potrebujeme danú funkciu interpolovať polynómom dostatočne vysokého stupňa. Tento postup nie je vždy najpohodlnejší a nemusí byť ani vhodný - pozri poznámku 4.3. Pri *zloženej kvadratúrnej metóde* sa vychádza z iného princípu. Daný interval $[a, b]$ sa rozdelí na dostatočný počet podintervalov a na nich sa funkcia f interpoluje polynómom nižšieho rádu. Teda f sa vlastne interpoluje na intervale $[a, b]$ splajnom.

Rozdeľme interval $[a, b]$ na m podintervalov $[x_j, x_{j+1}]$ rovnakej dĺžky $h = \frac{b-a}{m}$. Pre *lichobežníkovú metódu* teda dostaneme (pozri obrázok 8.1)

$$\begin{aligned}
\int_a^b f(x)dx &= \sum_{j=0}^{m-1} \int_{x_j}^{x_{j+1}} f(x)dx \\
&= \sum_{j=0}^{m-1} \left[\frac{b-a}{m} \, \frac{f(x_j) + f(x_{j+1})}{2} - \frac{h^3}{12} f^{(2)}(\xi_j) \right] \\
&= \frac{b-a}{m} \left[\frac{f(a)}{2} + f(x_1) + \ldots + f(x_{m-1}) + \frac{f(b)}{2} \right] - \frac{m\ h^3}{12} f^{(2)}(\xi) \\
&= \frac{b-a}{m} \left[\frac{f(a)}{2} + f(x_1) + \ldots + f(x_{m-1}) + \frac{f(b)}{2} \right] - \frac{(b-a)\ h^2}{12} f^{(2)}(\xi),
\end{aligned}$$

pre nejaké $\xi \in (a, b)$, pričom $\xi_j \in (x_j, x_{j+1})$. Poznamenajme, že ak $f \in C^2([a, b])$, tak také ξ existuje.

Uvažujme teraz Simpsonovu metódu. Na každom dielčom podintervale potrebujeme i jeho stred. Takže rozdeľme integračný interval $[a, b]$ na $2m$ podintervalov rovnakej dĺžky $h = \frac{b-a}{2m}$. Po úprave dostaneme

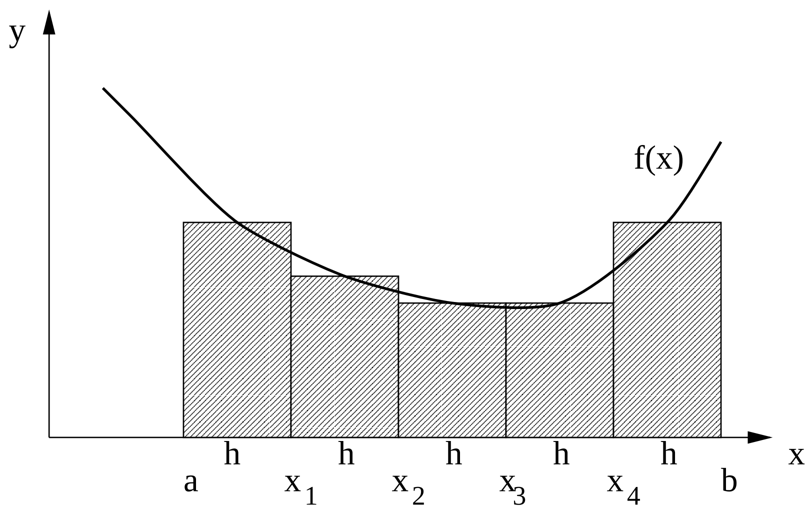
$$\int_a^b f(x)dx = \frac{b-a}{6m}\left[f(a)+f(b)+4\sum_{j=0}^{m-1} f(x_{2j+1})+2\sum_{j=1}^{m-1} f(x_{2j})\right] - \frac{h^4}{180}(b-a)f^{(4)}(\xi),$$

pre nejaké $\xi \in (a,b)$.

Pri odvodzovaní vzorca pre obdĺžnikovú metódu (pozri obrázok 8.2) opäť potrebujeme uvažovať stredy podintervalov, teda rozdeľme integračný interval $[a,b]$ na $2m$ podintervalov rovnakej dĺžky tak ako i pri Simpsonovej metóde. Výsledný vzorec bude

$$\int_a^b f(x)dx = \frac{b-a}{m}\sum_{j=0}^{m-1} f(x_{2j+1}) + \frac{h^2}{6}(b-a)f^{(2)}(\xi),$$

pre nejaké $\xi \in (a,b)$ a $h = \frac{b-a}{2m}$.



Obr. 8.2. Obdĺžniková metóda

8.2 Richardsonova extrapolácia

Predstavme si, že chceme určiť hodnotu funkcie $F(0)$. Pri integračných formulách $F(0)$ predstavuje presnú hodnotu určitého integrálu funkcie f. K dispozícii máme aproximačnú schému na výpočet $F(h)$ o presnosti $\mathcal{O}\left(h^k\right)$, teda

$$F(h) = F(0) + \sum_{j=k}^{\infty} a_j h^j. \tag{8.5}$$

Ukážme si ako sa dá zvýšiť presnosť uvažovanej schémy. Vypočítajme najprv $F(2h)$ a potom

$$
\begin{aligned}
2^k F(h) - F(2h) &= (2^k - 1)F(0) + \sum_{j=k}^{\infty} \left[2^k a_j h^j - a_j (2h)^j\right] \\
&= (2^k - 1)F(0) + \sum_{j=k}^{\infty} \left[2^k - 2^j\right] a_j h^j.
\end{aligned}
$$

Odtiaľ vyplýva

$$
F_e = \frac{2^k F(h) - F(2h)}{2^k - 1} = F(0) + \sum_{j=k+1}^{\infty} \frac{2^k - 2^j}{2^k - 1} a_j h^j. \tag{8.6}
$$

Všimnime si, že sumačný index j v poslednej sume začína hodnotou $k+1$. To znamená, že presnosť takto navrhnutej aproximácie bude $\mathcal{O}\left(h^{k+1}\right)$. Postup, ktorý sme si práve opísali sa nazýva Richardsonova extrapolácia. Vzťah (8.6) sme odvodili na základe asymptotického rozvoja (8.5). V niektorých prípadoch môže mať tento rozvoj špeciálny tvar, napr. obsahuje len párne alebo len nepárne mocniny h. Potom treba i toto špecifikum pri odvodzovaní vzorca pre Richardsonovu extrapoláciu zohľadniť. Uvažujme napr.

$$
F(h) = F(0) + \sum_{j=k}^{\infty} a_j h^{2j}. \tag{8.7}
$$

Potom analogicky dostaneme

$$
F_e = \frac{4^k F(h) - F(2h)}{4^k - 1} = F(0) + \sum_{j=k+1}^{\infty} \frac{4^k - 4^j}{4^k - 1} a_j h^{2j}. \tag{8.8}
$$

8.2.1 Rombergov kvadratúrny vzorec

Zovšeobecnime si Richardsonovu aproximáciu nasledujúcim spôsobom. Predstavme si, že sme urobili N výpočtov pre rôzne hodnoty h, t.j. pre rôzne počty podintervalov. Výpočty potom skombinujeme medzi sebou tak, aby prvých $N-1$ sčítancov v sume vo vzťahu (8.5) vypadlo. Aplikujme tento postup na lichobežníkovú metódu.

Nech $I_{0,k}$ sú približné hodnoty integrálu získané pomocou lichobežníkového pravidla s konštantnou dĺžkou podintervalov $\frac{b-a}{2^k}$. Aplikáciou (8.8) máme

$$
I_{m,k} = \frac{4^m I_{m-1,k+1} - I_{m-1,k}}{4^m - 1} \qquad k = 0, 1, \ldots\ . \tag{8.9}
$$

Postup vyzerá schematicky takto

$$
\left.
\begin{array}{lllll}
I_{0,0} & & & & \\
I_{0,1} & I_{1,0} & & & \\
I_{0,2} & I_{1,1} & I_{2,0} & & \\
\vdots & \vdots & \vdots & & \\
I_{0,m} & I_{1,m-1} & I_{2,m-2} & \ldots & I_{m,0}
\end{array}
\right\} \tag{8.10}
$$

V praxi to znamená, že najprv vypočítame približné hodnoty integrálu lichobežníkovým pravidlom s krokom $\frac{b-a}{2^k}$, t.j. prvý stĺpec v (8.10). Ďalšie stĺpce sa získajú pomocou (8.9). Takže z pôvodnej postupnosti $\{I_{0,k}\}_{k=0}^m$ takto získame "diagonálnu" postupnosť $\{I_{k,0}\}_{k=0}^m$, ktorá konverguje podstatne rýchlejšie. Prakticky si to ukážeme na nasledujúcom príklade.

Príklad 8.2 Rombergova schéma pre integrál

$$\int_1^2 \frac{dx}{x} = \log 2 \approx 0,6931471806$$

vyzerá takto

h_i	$I_{0,i}$	$I_{1,i}$	$I_{2,i}$	$I_{3,i}$	$I_{4,i}$
1	0,7500000000				
0,5	0,7083333333	0,6944444444			
0,25	0,6970238095	0,6932539683	0,6931746032		
0,125	0,6941218504	0,6931545307	0,6931479015	0,6931474776	
0,0625	0,6933912022	0,6931476528	0,6931471843	0,6931471831	**0,6931471819**

□

Cvičenie 8.1 Odvoďte integračnú formulku typu $\int_0^{2\pi} f(x)dx \approx Af(0) + Bf(\pi)$, ktorá má byť presná pre funkcie typu $f(x) = a + b\cos(x)$.

Cvičenie 8.2 Vypočítajte približnú hodnotu $\ln 2$ s presnosťou 10^{-4}. Pri výpočte použite zloženú kvadratúrnu metódu lichobežníkového typu.

Cvičenie 8.3 Integračná formulka

$$\int_0^1 f(x)dx \approx a_0 f(0) + a_1 f(0,5) + a_2 f(1)$$

má byť presná pre všetky polynómy najviac druhého stupňa. Určite koeficienty a_0, a_1, a_2.

Cvičenie 8.4 Odvoďte integračnú formulku tvaru

$$\int_a^b f(x)dx \approx \alpha[f(a) + f(b)],$$

ktorá je presná pre všetky polynómy najviac druhého stupňa.

Cvičenie 8.5 Koľko bodov potrebujete, aby

$$\int_0^1 \mathrm{e}^{-x^2}\, dx$$

ste vypočítali zloženým obdĺžnikovým pravidlom s presnosťou $\frac{1}{2}.10^{-4}$?

Kapitola 9

Obyčajné diferenciálne rovnice

Mnohé problémy z aplikovanej matematiky sa dajú opísat' pomocou obyčajných diferenciálnych rovníc (alebo systémov), ktoré opisujú časovú zmenu jednej alebo viacerých veličín. Zriedkavo sa dajú riešenia napísat' v analytickom tvare. Z tohto dôvodu je potrebné študovat' približné metódy pre riešenie obyčajných rovníc.

Na začiatku sa pre jednoduchost' a prehl'adnost' obmedzíme na jednu skalárnu rovnicu prvého rádu

$$y'(x) = f(x, y(x)), \tag{9.1}$$

ktorá spĺňa začiatočnú podmienku

$$y(x_0) = y_0. \tag{9.2}$$

O existencii a jednoznačnosti riešenia začiatočného problému hovorí nasledujúca veta.

Veta 9.1 *Nech funkcia f je spojitá v oboch svojich premenných x, y a nech f je globálne lipschitzovsky spojitá v y, t.j.*

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|, \qquad \forall x, y_1, y_2.$$

Potom rovnica (9.1) má jediné riešenie spĺňajúce (9.2).

V tejto kapitole si vysvetlíme niektoré jednokrokové metódy na riešenie obyčajných diferenciálnych rovníc. V závere si rozoberieme metódu sietí (diferenčnú metódu) a aplikujeme ju na riešenie okrajovej úlohy pre ODR druhého rádu.

Existujú i iné jednokrokové i viackrokové metódy s vyšším rádom presnosti výpočtu (napr. metódy Rungeho-Kutta). Tieto sú však pomerne komplikované, preto sa nimi nebudeme zaoberat'. Podrobnejšia analýza numerických metód sa dá nájst' napr. v [11], [14], [16].

9.1 Eulerova metóda napred

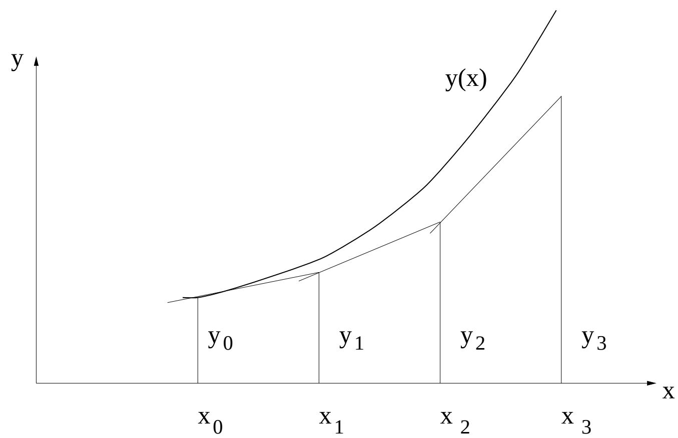
Označme uzly

$$x_n = x_0 + nh, \qquad n \in \mathbb{N},$$

pričom $h > 0$ označuje dĺžku kroku. Rovnica (9.1) v bode (x_0, y_0) určuje smernicu dotyčnice riešenia $y'(x_0) = f(x_0, y_0)$. Linearizáciou funkcie $y(x)$ dostaneme postupné aproximácie y_n pre exaktné hodnoty $y(x_n)$ na základe predpisu

$$y_{n+1} = y_n + hf(x_n, y_n), \qquad n = 0, 1, \dots \ . \tag{9.3}$$

Aproximácie sú schematicky zakreslené na obrázku 9.1. Táto metóda sa niekedy nazýva i mnohouholníková (polygónová) metóda.



Obr. 9.1. Eulerova metóda

Odhad chyby

Pri odvodení odhadu chyby budeme potrebovať túto lemu.

Lema 9.1 *Nech $\varepsilon_0 = 0$, $A = 1 + \alpha$, $\alpha > 0$ a pre všetky $n \in \mathbb{N}$ platí*

$$|\varepsilon_n| \le A|\varepsilon_{n-1}| + B.$$

Potom

$$|\varepsilon_n| \le \frac{A^n - 1}{A - 1} B \le \frac{e^{\alpha n} - 1}{\alpha} B.$$

DÔKAZ: Jednoduchým výpočtom dostaneme

$$\begin{aligned} |\varepsilon_n| &\le A|\varepsilon_{n-1}| + B \le A^2|\varepsilon_{n-2}| + B(1 + A) \\ &\le \ldots \le A^n|\varepsilon_0| + B\sum_{i=0}^{n-1} A^i \\ &= \frac{A^n - 1}{A - 1} B. \end{aligned}$$

Z rozvoja funkcie e^x do Taylorovho radu v bode 0 pre nejaké $\xi \in \mathbb{R}$ máme

$$e^x = 1 + x + e^{\xi}\,\frac{x^2}{2} \ge 1 + x, \qquad \forall x \in \mathbb{R}.$$

Takže

$$A^n = (1+\alpha)^n \le e^{\alpha n}$$

a nakoniec dostaneme

$$|\varepsilon_n| \le \frac{A^n - 1}{A-1} B \le \frac{e^{\alpha n} - 1}{\alpha} B. \qquad \square$$

Z rozvoja $y(x)$ do Taylorovho radu vyplýva pre nejaké $\xi_n \in (x_n, x_{n+1})$

$$\begin{aligned} y(x_{n+1}) &= y(x_n + h) = y(x_n) + hy'(x_n) + \frac{h^2}{2} y''(\xi_n) \\ &= y(x_n) + hf(x_n, y(x_n)) + \frac{h^2}{2} y''(\xi_n) \end{aligned} \tag{9.4}$$

Označme teraz $\varepsilon_n := y_n - y(x_n)$. Odčítaním (9.3) a (9.4) máme

$$\varepsilon_{n+1} = \varepsilon_n + h\left[f(x_n, y_n) - f(x_n, y(x_n))\right] - \frac{h^2}{2} y''(\xi_n).$$

Pomocou lipschitzovskej spojitosti funkcie f dostaneme

$$\begin{aligned} |\varepsilon_{n+1}| &\le |\varepsilon_n| + hL|y_n - y(x_n)| + \frac{h^2}{2} |y''(\xi_n)| \\ &\le |\varepsilon_n|(1 + Lh) + \frac{h^2}{2} \max_\xi |y''(\xi)| \\ &= A|\varepsilon_n| + B, \end{aligned}$$

pričom $A = 1 + Lh$, $B = \frac{h^2}{2} \max_\xi |y''(\xi)|$. Z lemy 9.1 vyplýva

$$|\varepsilon_n| \le B \frac{e^{nLh} - 1}{A - 1} = \frac{e^{(x_n - x_0)L} - 1}{2L} \max_\xi |y''(\xi)|\, h.$$

Takže ak $|x_n - x_0| \le C$ pre všetky $n \in \mathbb{N}$ a y'' je ohraničená, tak polygónová metóda je rádu $\mathcal{O}(h)$.

Vylepšená polygónová metóda

Predstavme si, že používame Eulerovu metódu napred s dĺžkou kroku h a sme v bode (x_k, y_k). Po jednom kroku dostaneme

$$y^1_{k+1} = y_k + hf(x_k, y_k). \tag{9.5}$$

Zmeňme teraz dĺžku kroku na $\frac{h}{2}$ a urobme dva kroky, t.j.

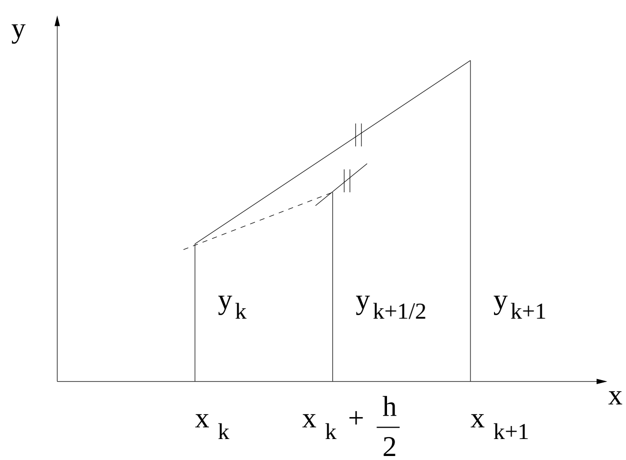
$$\begin{aligned} y^2_{k+\frac{1}{2}} &= y_k + \frac{h}{2} f(x_k, y_k), \\ y^2_{k+1} &= y^2_{k+\frac{1}{2}} + \frac{h}{2} f\left(x_k + \frac{h}{2}, y^2_{k+\frac{1}{2}}\right). \end{aligned} \tag{9.6}$$

Použitím Richardsonovej extrapolácie na y^2_{k+1} a y^1_{k+1} definujeme extrapolovanú hodnotu

$$\begin{aligned} y_{k+1} &= 2y^2_{k+1} - y^1_{k+1} = 2y^2_{k+\frac{1}{2}} + hf\left(x_k + \frac{h}{2}, y^2_{k+\frac{1}{2}}\right) - y_k - hf(x_k, y_k) \\ &= 2y_k + hf(x_k, y_k) + hf\left(x_k + \frac{h}{2}, y^2_{k+\frac{1}{2}}\right) - y_k - hf(x_k, y_k) \\ &= y_k + hf\left(x_k + \frac{h}{2}, y_k + \frac{h}{2} f(x_k, y_k)\right). \end{aligned}$$

Výsledok zapísaný algoritmicky je (pozri obrázok 9.2)

$$\begin{cases} k_1 &= f(x_k, y_k), \\ k_2 &= f\left(x_k + \frac{h}{2}, y_k + \frac{h}{2} k_1\right), \\ y_{k+1} &= y_k + hk_2. \end{cases} \tag{9.7}$$



Obr. 9.2. Vylepšená polygónová metóda

9.2 Eulerova spätná metóda

Pri polygónovej metóde sme nahradili deriváciu y' diferenciou napred, čiže $y'(x_n) \approx \dfrac{y_{n+1} - y_n}{h}$, a tak sme dostali explicitnú aproximačnú schému pre výpočet obyčajnej diferenciálnej rovnice prvého rádu. Pri Eulerovej spätnej metóde sa derivácia y' aproximuje diferenciou nazad $\left(y'(x_{n+1}) \approx \dfrac{y_{n+1} - y_n}{h}\right)$, teda

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}) \qquad n = 0, 1, \dots \ . \tag{9.8}$$

Táto schéma je implicitná vzhľadom na y_{n+1}, takže sa na každom kroku musí riešiť iteračným spôsobom.

Odhad chyby

Predpokladajme, že $y(x)$ je dostatočne hladká funkcia, ktorá sa dá rozvinúť do Taylorovho radu. Môžeme teda písať pre $\xi_n \in (x_n, x_{n+1})$

$$\begin{aligned} y(x_n) &= y(x_n+h) - hy'(x_n+h) + \frac{h^2}{2}y''(\xi_n) \\ &= y(x_{n+1}) - hf(x_{n+1}, y(x_{n+1})) + \frac{h^2}{2}y''(\xi_n) \ . \end{aligned}$$

Takže

$$y(x_{n+1}) = y(x_n) + hf(x_{n+1}, y(x_{n+1})) - \frac{h^2}{2}y''(\xi_n). \tag{9.9}$$

Označme teraz $\varepsilon_n := y_n - y(x_n)$. Odčítaním (9.8) a (9.9) dostaneme

$$\varepsilon_{n+1} = \varepsilon_n + h\left[f(x_{n+1}, y_{n+1}) - f(x_{n+1}, y(x_{n+1}))\right] + \frac{h^2}{2}y''(\xi_n).$$

Pretože f je globálne lipschitzovsky spojitá funkcia, tak

$$\begin{aligned} |\varepsilon_{n+1}| &\leq |\varepsilon_n| + hL|y_{n+1} - y(x_{n+1})| + \frac{h^2}{2}|y''(\xi_n)| \\ &\leq |\varepsilon_n| + hL|\varepsilon_{n+1}| + \frac{h^2}{2}\max_{\xi}|y''(\xi)| \ . \end{aligned}$$

Odtiaľ máme

$$|\varepsilon_{n+1}| \leq \frac{1}{1-Lh}|\varepsilon_n| + \frac{1}{1-Lh}\frac{h^2}{2}\max_{\xi}|y''(\xi)| = A|\varepsilon_n| + B,$$

pričom $A = 1+\alpha$, $\alpha = \dfrac{Lh}{1-Lh}$, $B = \dfrac{1}{1-Lh}\dfrac{h^2}{2}\max\limits_{\xi}|y''(\xi)|$.
Z lemy 9.1 vyplýva

$$\begin{aligned} |\varepsilon_n| &\leq B\frac{e^{n\alpha}-1}{A-1} = \frac{e^{\frac{Lnh}{1-Lh}}-1}{\frac{Lh}{1-Lh}}\,\frac{1}{1-Lh}\max_{\xi}|y''(\xi)|\frac{h^2}{2} \\ &= \frac{e^{\frac{(x_n-x_0)L}{1-Lh}}-1}{2L}\,\max_{\xi}|y''(\xi)|\ h. \end{aligned}$$

Takže, ak $|x_n - x_0| \leq C$ pre všetky $n \in \mathbb{N}$ a y'' je ohraničená, tak Eulerova spätná metóda je rádu $\mathcal{O}(h)$.

Príklad 9.1 Uvažujme diferenciálnu rovnicu so začiatočnou podmienkou tvaru

$$\begin{aligned} y'(t) + ay(t) &= f(t), \qquad t \in (0,T), \\ y(0) &= y_0, \end{aligned} \tag{9.10}$$

kde $a > 0$, $y_0 \in \mathbb{R}$, $f \in C([0,T])$. Ukážte platnosť apriórneho odhadu

$$\|y_n\| \leq C \qquad n = 1, \dots, N \ ,$$

kde $\|.\|$ je norma v priestore $L_2(0,T)$, t.j. $\|v\|^2 = (v,v) = \displaystyle\int_0^T v^2(t)dt$ a y_n je aproximácia riešenia v čase $t_n = n\tau$ ($\tau = \frac{T}{N}$, pre dosť veľké $N \in \mathbb{N}$), ktoré sme získali Eulerovou spätnou metódou, C je konštanta nezávislá od τ.

Riešenie: Aplikujeme Eulerovu spätnú metódu na (9.10), teda

$$\frac{y_i - y_{i-1}}{\tau} + a y_i = f_i \; . \tag{9.11}$$

Vynásobením (9.11) pomocou y_i, zintegrovaním cez $(0,T)$ a sčítaním pre $i = 1,\ldots,n$ dostaneme

$$\sum_{i=1}^{n}(y_i - y_{i-1}, y_i) + \sum_{i=1}^{n} \tau a(y_i, y_i) = \sum_{i=1}^{n} \tau (f_i, y_i) \; .$$

Pomocou identity

$$(q_i - q_{i-1}, q_i) = \frac{1}{2}\left\{(q_i, q_i) - (q_{i-1}, q_{i-1}) + (q_i - q_{i-1}, q_i - q_{i-1})\right\}$$

dostaneme

$$\frac{1}{2}\|y_n\|^2 - \frac{1}{2}\|y_0\|^2 + \frac{1}{2}\sum_{i=1}^{n}\|y_i - y_{i-1}\|^2 + a\tau\sum_{i=1}^{n}\|y_i\|^2 \le \frac{\tau}{2}\sum_{i=1}^{n}\|f_i\|^2 + \frac{\tau}{2}\sum_{i=1}^{n}\|y_i\|^2 ,$$

pričom sme v pravej strane použili Schwarzovu nerovnosť a vzťah $ab \le \frac{a^2+b^2}{2}$. Tretí a štvrtý člen na ľavej strane sú nezáporné, teda ich môžeme vynechať a nerovnosť sa pritom nezmení. Pretože funkcia f je ohraničená môžeme písať

$$\|y_n\|^2 \le C + \tau \sum_{i=1}^{n}\|y_i\|^2 ,$$

odkiaľ pre $0 < \tau < 1$ máme

$$\|y_n\|^2 \le \frac{C}{1-\tau} + \frac{1}{1-\tau}\sum_{i=1}^{n-1}\|y_i\|^2 \tau.$$

Použitím Gronwallovej lemy 3.2 dostaneme žiadaný apriórny odhad. □

9.3 Crankova-Nicolsonova metóda

V tejto časti si ukážeme ako sa jednoduchou kombináciou spätnej Eulerovej a polygónovej metódy dá zvýšiť rád konvergencie.

Uvažujme Crankovu-Nicolsonovu schému

$$y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, y_{n+1})\right], \qquad n = 0, 1, \ldots \; . \tag{9.12}$$

Ide tu opäť o implicitnú metódu. Rozvojom funkcie y do Taylorovho radu dostaneme

$$\begin{aligned} y(x_{n+1}) &= y(x_n) + h y'(x_n) + \tfrac{h^2}{2} y''(\xi_n), \\ y(x_n) &= y(x_{n+1}) - h y'(x_{n+1}) + \tfrac{h^2}{2} y''(\eta_n), \end{aligned}$$

pre nejaké $\xi_n, \eta_n \in (x_n, x_{n+1})$. Odčítaním oboch rovníc máme

$$\begin{aligned} y(x_{n+1}) - y(x_n) &= y(x_n) - y(x_{n+1}) + h\left[f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))\right] \\ &+\frac{h^2}{2}(y''(\xi_n) - y''(\eta_n)) \\ &= y(x_n) - y(x_{n+1}) + h\left[f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))\right] \\ &+\frac{h^2}{2}y'''(\theta_n)(\xi_n - \eta_n), \end{aligned}$$

pre θ_n medzi ξ_n a η_n. Odtiaľ vyplýva

$$y(x_{n+1}) = y(x_n) + \frac{h}{2}\left[f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))\right] + \frac{h^2}{4}y'''(\theta_n)(\xi_n - \eta_n). \tag{9.13}$$

Odčítaním (9.12) a (9.13) dostaneme

$$\begin{aligned} y_{n+1} - y(x_{n+1}) &= y_n - y(x_n) + \frac{h}{2}\left[f(x_n, y_n) - f(x_n, y(x_n))\right] \\ &+\frac{h}{2}\left[f(x_{n+1}, y_{n+1}) - f(x_{n+1}, y(x_{n+1}))\right] - \frac{h^2}{4}y'''(\theta_n)(\xi_n - \eta_n), \end{aligned}$$

teda pomocou lipschitzovskosti funkcie f odhadneme

$$|\varepsilon_{n+1}| \le |\varepsilon_n|\left(1 + \frac{Lh}{2}\right) + |\varepsilon_{n+1}|\frac{Lh}{2} + \frac{h^3}{4}\max_\theta |y'''(\theta)|.$$

Takže

$$|\varepsilon_{n+1}| \le A|\varepsilon_n| + B,$$

pričom $A = 1 + \alpha$, $\alpha = \dfrac{Lh}{1 - \frac{Lh}{2}}$, $B = \dfrac{1}{1 - \frac{Lh}{2}}\dfrac{h^3}{4}\max\limits_\theta |y'''(\theta)|$.
Z lemy 9.1 vyplýva

$$|\varepsilon_n| \quad \le B\frac{e^{n\alpha} - 1}{A - 1} = \frac{e^{\frac{(x_n - x_0)L}{1 - \frac{Lh}{2}}} - 1}{4L} \max_\theta |y'''(\theta)|\ h^2.$$

Takže, ak $|x_n - x_0| \le C$ pre všetky $n \in \mathbb{N}$ a y''' je ohraničená, tak presnosť Crankovej-Nicolsonovej metódy je $\mathcal{O}\left(h^2\right)$.

9.4 Okrajová úloha pre ODR 2. rádu

Uvažujme nasledujúcu okrajovú úlohu

$$\begin{aligned} -u''(x) + c(x)u(x) &= f(x), \qquad 0 < x < l, \\ u(0) = u(l) &= 0. \end{aligned} \tag{9.14}$$

Predpokladajme, že funkcie c, f sú spojité na intervale $[0, l]$ a navyše $c(x) \ge 0$ na celom $[0, l]$. Za týchto predpokladov má problém (9.14) jediné riešenie (pozri [6]). Označme diferenciálny operátor $A = -\frac{d^2}{dx^2} + c(x)$. Oblasť definície operátora $D(A)$ pozostáva zo všetkých funkcií z $X = C([0,l])$, ktoré sú dvakrát spojite diferencovateľné a spĺňajú okrajové podmienky z (9.14).

Rozdeľme interval $[0,l]$ na n rovnakých podintervalov dĺžky $h = \frac{l}{n}$ a s uzlami $x_k = kh$ pre $k = 0,\dots,n$. Pre $u \in X$ definujme operátor zúženia $T_n u(x) = \{u(x_k)\}_{k=0}^n$ (na okrajoch sú fixné hodnoty $u(x_0) = u(x_n) = 0$). Zaveďme sieťový priestor X_n vektorov dĺžky $n+1$ (s prvou a poslednou nulovou zložkou) so skalárnym súčinom

$$(\boldsymbol{u},\boldsymbol{v}) = \frac{l}{n}\sum_{k=1}^{n-1} u_k v_k$$

pre vektory $\boldsymbol{v} = (v_0,\dots,v_n)^T$, $\boldsymbol{u} = (u_0,\dots,u_n)^T$, $u_0 = u_n = v_0 = v_n = 0$. Zodpovedajúca sférická norma bude

$$\|\boldsymbol{u}\| = \sqrt{\frac{l}{n}\sum_{k=1}^{n-1} u_k^2}.$$

Predpokladajme, že riešenie $u(x)$ úlohy (9.14) je 4-krát spojite diferencovateľné v $(0,l)$, pričom

$$\gamma = \sup_{x\in(0,l)} |u^{(4)}(x)| < \infty. \tag{9.15}$$

Z Taylorovho rozvoja vyplýva

$$\begin{aligned} u(x_{k+1}) &= u(x_k) + hu^{(1)}(x_k) + \frac{h^2}{2}u^{(2)}(x_k) + \frac{h^3}{6}u^{(3)}(x_k) + \frac{h^4}{24}u^{(4)}(\theta_k), \\ u(x_{k-1}) &= u(x_k) - hu^{(1)}(x_k) + \frac{h^2}{2}u^{(2)}(x_k) - \frac{h^3}{6}u^{(3)}(x_k) + \frac{h^4}{24}u^{(4)}(\eta_k), \end{aligned}$$

pre nejaké $\theta_k \in (x_k, x_{k+1})$ a $\eta_k \in (x_{k-1}, x_k)$. Odtiaľ vidíme, že aproximácia u'' centrálnou diferenciou je rádu $\mathcal{O}(h^2)$, teda

$$u''(x_k) = \frac{u(x_{k+1}) - 2u(x_k) + u(x_{k-1})}{h^2} + \mathcal{O}\left(h^2\right).$$

Aproximujme (9.14)

$$\begin{aligned} -\frac{u_{k+1} - 2u_k + u_{k-1}}{h^2} + c_k u_k &= f_k, \qquad k = 1,\dots,n-1 \\ u_0 = u_n &= 0, \end{aligned} \tag{9.16}$$

pričom $\boldsymbol{T}_n(f) = \{f(x_k)\}_{k=1}^{n-1} = \{f_k\}_{k=1}^{n-1} = \boldsymbol{f}_n$ a $\boldsymbol{T}_n(c) = \{c(x_k)\}_{k=1}^{n-1} = \{c_k\}_{k=1}^{n-1}$. Diferenčná schéma (9.16) v maticovom tvare vyzerá takto

$$\boldsymbol{A}_n \boldsymbol{u}_n = \boldsymbol{f}_n, \tag{9.17}$$

pričom

$$
\boldsymbol{A}_n = \begin{pmatrix}
\frac{2}{h^2}+c_1 & -\frac{1}{h^2} & 0 & \dots & & & & 0 \\
-\frac{1}{h^2} & \frac{2}{h^2}+c_2 & -\frac{1}{h^2} & 0 & \dots & & & 0 \\
0 & -\frac{1}{h^2} & \frac{2}{h^2}+c_3 & -\frac{1}{h^2} & 0 & \dots & & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & \dots & & 0 & -\frac{1}{h^2} & \frac{2}{h^2}+c_{n-3} & -\frac{1}{h^2} & 0 \\
0 & \dots & & & 0 & -\frac{1}{h^2} & \frac{2}{h^2}+c_{n-2} & -\frac{1}{h^2} \\
0 & \dots & & & & 0 & -\frac{1}{h^2} & \frac{2}{h^2}+c_{n-1}
\end{pmatrix}.
$$

Uvažujme ľubovoľný prvok $\boldsymbol{v} \in X_n$. Pomocou Cauchyho nerovnosti odhadneme

$$
\begin{aligned}
\|\boldsymbol{v}\|^2 &= h\sum_{i=1}^{n-1}|v_i|^2 && = h\sum_{i=1}^{n-1}\left|\sum_{k=1}^{i}(v_k - v_{k-1})\right|^2 \\
&\le h\sum_{i=1}^{n-1} i \sum_{k=1}^{i}|v_k - v_{k-1}|^2 && \le h\sum_{i=1}^{n-1} i \sum_{k=1}^{n}|v_k - v_{k-1}|^2 \\
&= \frac{n(n-1)h}{2}\sum_{k=1}^{n}|v_k - v_{k-1}|^2 && \le \frac{l^2}{2h}\sum_{k=1}^{n}|v_k - v_{k-1}|^2 .
\end{aligned} \tag{9.18}
$$

Jednoduchým výpočtom sa môžeme presvedčiť, že

$$
\sum_{k=1}^{n}|v_k - v_{k-1}|^2 = -\sum_{k=1}^{n-1}(v_{k+1} - 2v_k + v_{k-1})v_k. \tag{9.19}
$$

Ďalej

$$
(\boldsymbol{A}_n\boldsymbol{v}, \boldsymbol{v}) = h\sum_{k=1}^{n-1} c_k v_k^2 - \frac{1}{h}\sum_{k=1}^{n-1}(v_{k+1} - 2v_k + v_{k-1})v_k. \tag{9.20}
$$

Kombináciou (9.18), (9.19) a (9.20) spolu s nezápornosťou funkcie c máme

$$
\frac{2}{l^2}\|\boldsymbol{v}\|^2 \le (\boldsymbol{A}_n\boldsymbol{v}, \boldsymbol{v}),
$$

teda

$$
\frac{2}{l^2}\|\boldsymbol{v}\| \le \|\boldsymbol{A}_n\boldsymbol{v}\|. \tag{9.21}
$$

Zvoľme teraz $\boldsymbol{v} = \boldsymbol{u}_n - \boldsymbol{T}_n u$. Platí

$$
\begin{aligned}
\frac{2}{l^2}\ \|\boldsymbol{u}_n - \boldsymbol{T}_n u\| &\le \|\boldsymbol{A}_n\boldsymbol{u}_n - \boldsymbol{A}_n\boldsymbol{T}_n u\| \\
&\le \|\boldsymbol{A}_n\boldsymbol{u}_n - \boldsymbol{T}_n Au\| + \|\boldsymbol{T}_n Au - \boldsymbol{A}_n\boldsymbol{T}_n u\| \\
&= \|\boldsymbol{T}_n Au - \boldsymbol{A}_n\boldsymbol{T}_n u\|.
\end{aligned}
$$

Posledný výraz odhadneme takto

$$\begin{aligned}\|\boldsymbol{T}_n Au - \boldsymbol{A}_n\boldsymbol{T}_n u\|^2 &= h\sum_{k=1}^{n-1}\left|u''(x_k) - \frac{u(x_{k+1}) - 2u(x_k) + u(x_{k-1})}{h^2}\right|^2 \\ &= \mathcal{O}\left(h^4\right).\end{aligned}$$

Týmto sme dokázali

$$\|\boldsymbol{u}_n - \boldsymbol{T}_n u\| = O\left(\frac{1}{n^2}\right) = \mathcal{O}\left(h^2\right).$$

Pomocou nerovnosti o ekvivalentnosti noriem

$$\sqrt{\frac{l}{m}}\max_{1\le i\le m}|v_i| \le \sqrt{\frac{l}{m}\sum_{i=1}^{m}|v_i|^2} \le \sqrt{l}\max_{1\le i\le m}|v_i| \tag{9.22}$$

dostaneme tvrdenie o bodovej konvergencii siet'ovej metódy

$$\max_{1\le i\le n-1}|u_i - u(x_i)| = O\left(\frac{1}{n^2}\right) = \mathcal{O}\left(h^2\right).$$

Cvičenie 9.1 Dokážte platnost' (9.22).

Cvičenie 9.2 Vo vzt'ahu (9.11) označme $\delta y_i = \dfrac{y_i - y_{i-1}}{\tau}$. Ukážte, že existuje také $C > 0$, že pre dostatočne malé τ a pre $1 \le n \le N$ platí

$$\sum_{i=1}^{n}\|\delta y_i\|^2\tau \le C.$$

Cvičenie 9.3 Uvažujme približnú schému

$$\frac{y_i - y_{i-1}}{\tau} + ay_i = f(y_{i-1})$$

pre $1 \le i \le N$ a pre dané $a > 0$. Podobne ako v (9.11) dokážte, že ak funkcia f je lipschitzovsky spojitá, potom vzt'ah

$$\|y_i\|^2 \le C$$

platí pre všetky $1 \le i \le N$.

Cvičenie 9.4 Riešte okrajovú úlohu pre obyčajnú diferenciálnu rovnicu druhého rádu

$$\begin{aligned}-u''(x) &= 2 \qquad \text{v intervale } (0,3) \\ u(0) &= -9 \\ -u'(3) &= 0\end{aligned}$$

na danej sieti $x_i = i$ pre $i = 0,1,2,3$.

Cvičenie 9.5 Riešte okrajovú úlohu pre obyčajnú diferenciálnu rovnicu druhého rádu

$$\begin{aligned}-u''(x) &= 2 \qquad \text{v intervale } (0,4) \\ u(0) &= 0 \\ u(4) &= 8\end{aligned}$$

na danej sieti $x_i = i$ pre $i = 0,1,2,3,4$.

Kapitola 10

Parciálne diferenciálne rovnice

Viaceré stavy vo fyzike, chémii či biológii sa dajú opísať pomocou funkcií viacerých premenných. Ich zmeny v priestore alebo čase sa opisujú pomocou parciálnych diferenciálnych rovníc (PDR), ktoré vyjadrujú napr. zákon zachovania hmoty, energie, hybnosti atď. Rôznorodosť takto získaných parciálnych diferenciálnych rovníc alebo systémov je príčinou vývoja špeciálnych numerických metód pre ich riešenie.

Všeobecná lineárna PDR dvoch premenných má tvar

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + Fu = H,$$

pričom koeficienty A, B, C, D, E, F, H môžu byť i funkciami premenných x, y. Analogicky ako aj pri klasifikácii kužeľosečiek

$$Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F = 0$$

rozdeľujeme PDR do troch základných tried:

1. PDR ELIPTICKÉ:]
   ak $AC - B^2 > 0$ pre všetky (x, y) v danej oblasti Ω. Kánonický tvar[1] je

$$u_{xx} + u_{yy} + \alpha u_x + \beta u_y + \gamma u = f(x, y).$$

Známe príklady sú

- Laplaceova rovnica $\quad -u_{xx} - u_{yy} = 0,$
- Poissonova rovnica $\quad -u_{xx} - u_{yy} = f(x, y).$

2. PDR HYPERBOLICKÉ:
   ak $AC - B^2 < 0$ pre všetky $(x, y) \in \Omega$. Kánonický tvar je

$$u_{xy} + \alpha u_x + \beta u_y + \gamma u = f(x, y).$$

Známy príklad je

- vlnová rovnica $\quad -u_{xx} + u_{yy} = f(x, y).$

[1] Podrobnosti o transformácii PDR na jej kánonický tvar sú uvedené napr. v [1].

3. PDR PARABOLICKÉ:
ak $AC - B^2 = 0$ pre všetky $(x, y) \in \Omega$. Kánonický tvar je

$$u_{xx} + \alpha u_x + \beta u_y + \gamma u = f(x, y).$$

Známy príklad je

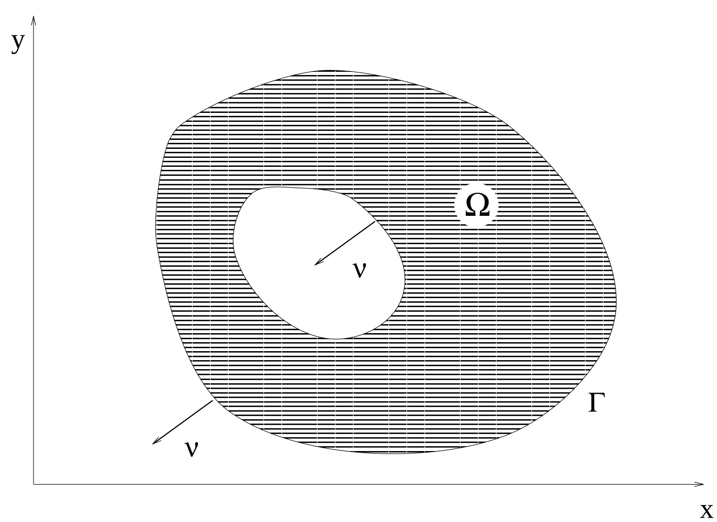
- difúzna rovnica $\qquad u_y - u_{xx} = f(x, y).$

Okrajové podmienky
V mnohých aplikáciách je uvažovaná oblasť Ω ohraničená a je známe správanie sa riešenia PDR na hranici oblasti $\Gamma = \partial\Omega$. Rozoznávame tri základné typy hraničných podmienok:

DIRICHLETOVA PODMIENKA: Daná je hodnota riešenia $u = \phi$ na $\Gamma_D \subset \Gamma$.

NEUMANOVA PODMIENKA: Daná je normálová zložka toku (cez hranicu $\Gamma_N \subset \Gamma$) $-D\nabla u \cdot \boldsymbol{\nu} = \gamma$, pričom $\boldsymbol{\nu}$ je vektor vonkajšej normály ku hranici v danom bode (pozri obrázok 10.1).

ROBINOVA PODMIENKA: Normálová zložka toku cez hranicu závisí od hodnoty riešenia v danom bode, t.j. $-D\nabla u \cdot \boldsymbol{\nu} + \alpha u = \beta$ na $\Gamma_R \subset \Gamma$.



Obr. 10.1. Oblasť, jej hranica a vektor vonkajšej normály

Všetky tri typy podmienok sa môžu vyskytovať v danej úlohe. Vtedy sú však jednotlivé časti navzájom disjunktné, t.j. $\Gamma_D \cap \Gamma_N = \Gamma_D \cap \Gamma_R = \Gamma_R \cap \Gamma_N = \emptyset$ a pokrývajú celú hranicu $\Gamma = \Gamma_D \cup \Gamma_N \cup \Gamma_R$.

Niekedy sa Dirichletova, Neumanova a Robinova podmienka označujú ako podmienky 1., 2. resp. 3. typu alebo tiež sa pre Robinovu podmienku môžeme v literatúre stretnúť s názvom Newtonova podmienka.

10.1 PDR eliptického typu

Eliptické diferenciálne rovnice zohrávajú významnú úlohu pri štúdii PDR. Podmienka eliptičnosti pre diferenciálny operátor 2. rádu v $\mathbb{R}^n$ tvaru

$$Au = -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) + cu$$

je charakterizovaná nerovnosťou

$$\sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j \geq C_0 \sum_{i=1}^{n} \xi_i^2,$$

ktorá musí byť splnená pre všetky $(\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$, pričom $C_0 > 0$.

Pojem eliptičnosti nie je viazaný na rovnice druhého rádu a dá sa zovšeobecniť i na rovnice vyšších rádov. Typickým príkladom eliptickej diferenciálnej rovnice štvrtého rádu je biharmonická rovnica v n premenných

$$\left(\sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}\right)^2 u(x_1, \ldots, x_n) = f(x_1, \ldots, x_n).$$

Najjednoduchšou približnou metódou na riešenie PDR je *metóda sietí.* Táto je založená na aproximácii derivácií diferenčnými podielmi. Druhá skupina numerických metód na riešenie PDR je založená na variačnom princípe. Z nich najdôležitejšia je metóda konečných prvkov, táto však presahuje rámec tejto publikácie, a preto sa ňou nebudeme zaoberať.

Dvojrozmerná eliptická úloha

Uvažujme obdĺžnikovú oblasť $\Omega = (0, a) \times (0, b)$. Chceme riešiť Dirichletovu úlohu

$$\begin{aligned} -D\left(\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2}\right) &= f(x,y) && \text{v } \Omega, \\ u &= \phi && \text{na } \partial\Omega, \end{aligned} \tag{10.1}$$

pričom $D \in \mathbb{R}_+$. Pravá strana f i okrajová podmienka ϕ sú dostatočne hladké funkcie svojich premenných. Z teórie lineárnych eliptických rovníc vyplýva existencia i jednoznačnosť riešenia problému (10.1) (pozri [6]). Predpokladajme, že riešenie úlohy u má spojité štvrté derivácie a

$$\gamma = \sup_{(x,y)\in\overline{\Omega}} \left(\left|\frac{\partial^4 u(x,y)}{\partial x^4}\right| + \left|\frac{\partial^4 u(x,y)}{\partial y^4}\right|\right) < \infty.$$

Nahraďme druhé parciálne derivácie centrálnymi diferenciami pre dve premenné, teda

$$\begin{aligned} \frac{\partial^2 u(x,y)}{\partial x^2} &= \frac{u(x+\Delta x, y) - 2u(x,y) + u(x-\Delta x, y)}{(\Delta x)^2} + \mathcal{O}\left((\Delta x)^2\right), \\ \frac{\partial^2 u(x,y)}{\partial y^2} &= \frac{u(x, y+\Delta y) - 2u(x,y) + u(x, y-\Delta y)}{(\Delta y)^2} + \mathcal{O}\left((\Delta y)^2\right). \end{aligned}$$

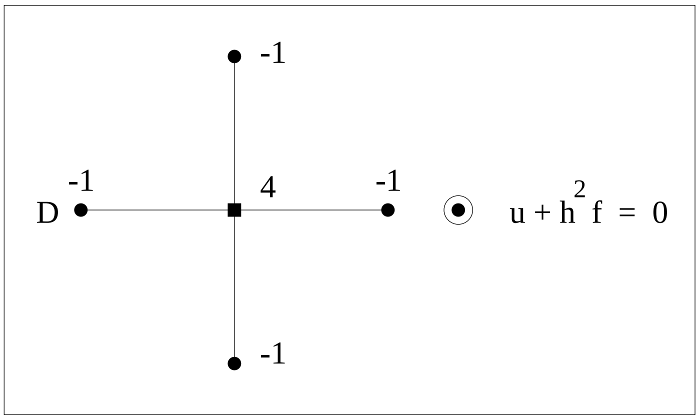
Rozdeľme interval $[0, a]$ rovnomerne na L podintervalov dĺžky $h = \Delta x = \frac{a}{L}$. Uzlové body označme $x_l = lh$ pre $l = 0, \ldots, L$. Podobne rozdeľme interval $[0, b]$ na M rovnakých podintervalov o dĺžke[2] $h = \Delta y = \frac{b}{M}$. Uzlové body budú $y_m = mh$ pre $m = 0, \ldots, M$. Rovnica (10.1) v typickom uzle siete (x_l, y_m) bude mať tvar

$$-D\left(\frac{\partial^2 u(x_l, y_m)}{\partial x^2} + \frac{\partial^2 u(x_l, y_m)}{\partial y^2}\right) = f(x_l, y_m) = f_{l,m}. \tag{10.2}$$

Diferenčná aproximačná schéma bude

$$-D\left(\frac{u_{l+1,m} - 2u_{l,m} + u_{l-1,m}}{h^2} + \frac{u_{l,m+1} - 2u_{l,m} + u_{l,m-1}}{h^2}\right) = f_{l,m}, \tag{10.3}$$

pre $l = 1, \ldots, L-1$ a $m = 1, \ldots, M-1$. Táto schéma sa dá symbolicky vyjadriť v operátorovom tvare, ktorý je zachytený na obrázku 10.2.



Obr. 10.2. Operátorová forma diferenčnej schémy

V tomto sú už zahrnuté i okrajové podmienky, takže musí platiť

$$\begin{aligned} u_{0,m} &= \phi(0, y_m), & u_{L,m} &= \phi(a, y_m) \\ u_{l,0} &= \phi(x_l, 0), & u_{l,M} &= \phi(x_l, b). \end{aligned}$$

Spolu máme teda $(L-1) \times (M-1)$ vnútorných bodov, t.j. neznámych. Hraničné hodnoty sú nahradené hodnotami okrajovej podmienky v príslušnom bode (pozri obrázok 10.3).

Označme

$$\boldsymbol{u}_l^T = (u_{l,1}, u_{l,2}, \ldots, u_{l,M-1})$$

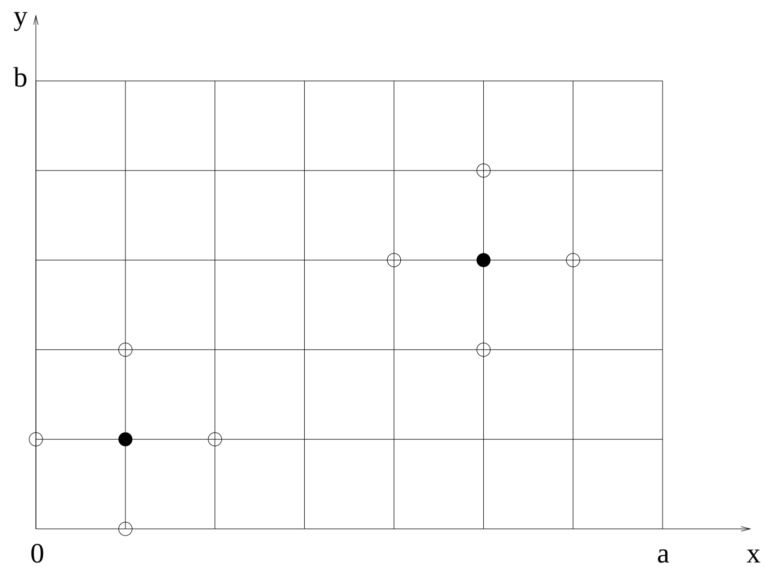
a

$$\boldsymbol{u}^T = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \ldots, \boldsymbol{u}_{L-1}).$$

Ďalej budeme potrebovať maticu $\overline{\boldsymbol{K}}$ dimenzie $(M-1) \times (M-1)$ tvaru

$$\overline{\boldsymbol{K}} = \begin{pmatrix} 4 & -1 & 0 & \ldots & & & & 0 \\ -1 & 4 & -1 & 0 & \ldots & & & 0 \\ 0 & -1 & 4 & -1 & 0 & \ldots & & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & & & 0 & -1 & 4 & -1 \\ 0 & \ldots & & & & 0 & -1 & 4 \end{pmatrix}$$

[2]Nie je nutné aby sme uvažovali ten istý krok v smere osi x aj osi y. Ak $\Delta x = \Delta y$, tak budú vzorce prehľadnejšie.



Obr. 10.3. Dve rôzne polohy šablóny na sieti. Znak • označuje stred šablóny a ∘ jej konce

a jednotkovú maticu $\boldsymbol{I}$ dimenzie $(M-1) \times (M-1)$.
Systém rovníc (10.3) sa dá zapísať v maticovom tvare

$$\boldsymbol{K u} = \boldsymbol{F}, \tag{10.4}$$

kde matica $\boldsymbol{K}$ je dimenzie $(L-1) \times (L-1)$ tvaru

$$\boldsymbol{K} = \begin{pmatrix} \overline{\boldsymbol{K}} & -\boldsymbol{I} & 0 & \dots & & & & 0 \\ -\boldsymbol{I} & \overline{\boldsymbol{K}} & -\boldsymbol{I} & 0 & \dots & & & 0 \\ 0 & -\boldsymbol{I} & \overline{\boldsymbol{K}} & -\boldsymbol{I} & 0 & \dots & & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & & & 0 & -\boldsymbol{I} & \overline{\boldsymbol{K}} & -\boldsymbol{I} \\ 0 & \dots & & & & 0 & -\boldsymbol{I} & \overline{\boldsymbol{K}} \end{pmatrix},$$

$$\boldsymbol{F}^{\boldsymbol{T}} = (\boldsymbol{F}_1, \boldsymbol{F}_2, \dots, \boldsymbol{F}_{L-1}),$$

$$\boldsymbol{F}_{\boldsymbol{l}}^{\boldsymbol{T}} = (F_{l,1}, \dots, F_{l,M-1}),$$

pričom

$$F_{l,m} = \frac{h^2}{D} f_{l,m} + \phi(x_l, 0)\delta_{m,1} + \phi(x_l, b)\delta_{m,M-1} + \phi(0, x_m)\delta_{l,1} + \phi(a, x_m)\delta_{l,L-1}.$$

Takto sme dostali formálne podobný systém s (9.17). Všimnime si, že pôvodne trojdiagonálna matica v (9.17) sa zmenila na blokovo trojdiagonálnu v (10.4). Táto je pozitívne definitná, a preto sa dá sústava rovníc (10.4) riešiť štandartnými metódami. Takto sa získa približné riešenie Poissonovej rovnice (10.1). Podobne ako aj v jednorozmernom prípade sa i tu dá odvodiť odhad chyby aproximácie diferenčnou schémou

$$\max_{i,j} |u_{i,j} - u(x_i, y_j)| = \mathcal{O}\left(h^2\right).$$

Neumanova okrajová podmienka

Zmeňme pôvodný problém (10.1) na nasledujúci. Chceme riešiť zmiešanú okrajovú úlohu

$$\begin{aligned} -D\left(\frac{\partial^2 u(x,y)}{\partial x^2}+\frac{\partial^2 u(x,y)}{\partial y^2}\right) &= f(x,y) && \text{v } \Omega=(0,a)\times(0,b), \\ u(x,y) &= \phi(x,y) && \text{na } \partial\Omega \text{ ak } y\neq 0, \\ -D\nabla u(x,y)\cdot\boldsymbol{\nu} &= 0 && \text{na } \partial\Omega \text{ ak } y=0. \end{aligned} \tag{10.5}$$

Teraz sú neznáme hodnoty riešenia i na časti hranice $y=0$. Podmienku nulového toku $-D\nabla u(x,y)\cdot\boldsymbol{\nu}=0$ cez túto časť hranice si namodelujeme takto. Predstavme si, že sme pomyslene predĺžili $u_{l,m}$ aj pre záporné indexy m. Ak má byť normálová zložka toku cez os $y=0$ nulová, tak potrebujeme podmienku symetrie riešenia vzhľadom na os $y=0$

$$u_{l,-1}=u_{l,1} \qquad \text{pre } l=1,\dots,L-1. \tag{10.6}$$

Diferenčná aproximačná schéma bude

$$-D\left(\frac{u_{l+1,m}-2u_{l,m}+u_{l-1,m}}{h^2}+\frac{u_{l,m+1}-2u_{l,m}+u_{l,m-1}}{h^2}\right)=f_{l,m}, \tag{10.7}$$

pre $l=1,\dots,L-1$ a $m=0,\dots,M-1$. V tomto sú už zahrnuté i okrajové podmienky, takže musí platiť

$$\begin{aligned} u_{0,m} &= \phi(0,y_m), & u_{L,m} &= \phi(a,y_m), \\ u_{l,M} &= \phi(x_l,b). \end{aligned}$$

Takže spolu máme $(L-1)\times M$ neznámych a taký istý počet algebraických rovníc.

10.2 Jednorozmerná parabolická úloha

V tejto časti sa budeme zaoberať jednorozmernou úlohou pre vedenie tepla

$$\begin{aligned} \frac{\partial u(t,x)}{\partial t}-D\frac{\partial^2 u(t,x)}{\partial x^2} &= f(t,x), && (t,x)\in(0,T)\times(0,l), \\ u(t,0)=u(t,l) &= 0 && \text{pre } t>0, \\ u(0,x) &= \phi(x), && \text{pre } x\in(0,l), \end{aligned} \tag{10.8}$$

pričom $T,D\in\mathbb{R}_+$ a f,ϕ sú dostatočne hladké funkcie svojich premenných. Ďalej predpokladáme, že začiatočné i okrajové podmienky sú navzájom kompatibilné, t.j. $\phi(0)=\phi(l)=0$. Z teórie lineárnych parabolických rovníc potom vyplýva existencia i jednoznačnosť *klasického riešenia* problému (10.8).

Implicitná diferenčná schéma

Rozdeľme časový interval $[0,T]$ na n rovnakých podintervalov dĺžky $\tau=\frac{T}{n}$ s časovými krokmi $t_i=i\tau$ pre $i=1,\dots,n$. Priestorový interval $[0,l]$ rozdeľme na m podintervalov $[x_{j-1},x_j]$ pre $j=1,\dots,m$ rovnakej dĺžky $h=\frac{l}{m}$. Teda $x_j=jh$. Ak nahradíme časovú deriváciu diferenciou nazad a priestorovú deriváciu druhého rádu centrálnou diferenciou, dostaneme implicitnú diferenčnú schému

$$\frac{u_j^i-u_j^{i-1}}{\tau}-D\frac{u_{j+1}^i-2u_j^i+u_{j-1}^i}{h^2}=f_j^i=f(t_i,x_j) \tag{10.9}$$

pre $i = 1, \dots, n$, $j = 1, \dots, m-1$. Okrajové podmienky budú

$$u_0^i = u_m^i = 0 \qquad \text{pre } i = 1, \dots, n. \tag{10.10}$$

Začiatočná podmienka znamená

$$u_j^0 = \phi_j = \phi(x_j) \qquad \text{pre } j = 0, \dots, m. \tag{10.11}$$

Rovnicu (10.9) si môžeme prepísať takto

$$-D\frac{u_{j+1}^i - 2u_j^i + u_{j-1}^i}{h^2} + \frac{u_j^i}{\tau} = \frac{u_j^{i-1}}{\tau} + f_j^i.$$

To znamená, že pre pevné i sme dostali analógiu diferenčnej schémy (9.16) pre obyčajnú diferenciálnu rovnicu druhého rádu, ktorú sme už študovali. Takže na každom časovom reze i máme zaručenú existenciu i jednoznačnosť riešenia u_j^i pre $j = 1, \dots, m-1$.

Stabilita aproximačného riešenia

Cieľom tejto časti je získať apriórny odhad riešenia u_j^i (vo vhodnej norme) nezávislý od i, j, t.j. od času i miesta. Upravme si preto (10.9) takto

$$u_j^i\left(1 + \frac{2D\tau}{h^2}\right) = \frac{D\tau}{h^2}\left(u_{j-1}^i + u_{j+1}^i\right) + \tau f_j^i + u_j^{i-1}.$$

Odtiaľ pre absolútne hodnoty dostaneme

$$\begin{aligned} |u_j^i|\left(1 + \tfrac{2D\tau}{h^2}\right) &\le \frac{D\tau}{h^2}\left(|u_{j-1}^i| + |u_{j+1}^i|\right) + \tau|f_j^i| + |u_j^{i-1}| \\ &\le \frac{2D\tau}{h^2}\max_j |u_j^i| + \tau\max_j |f_j^i| + \max_j |u_j^{i-1}|. \end{aligned}$$

Takže platí

$$\max_j |u_j^i|\left(1 + \frac{2D\tau}{h^2}\right) \le \frac{2D\tau}{h^2}\max_j |u_j^i| + \tau\max_j |f_j^i| + \max_j |u_j^{i-1}|,$$

teda

$$\begin{aligned} \max_j |u_j^i| &\le \tau\max_j |f_j^i| + \max_j |u_j^{i-1}| \\ &\le \tau\left(\max_j |f_j^i| + \max_j |f_j^{i-1}|\right) + \max_j |u_j^{i-2}| \\ &\le \dots \\ &\le \max_j |\phi_j| + \sum_{k=1}^{i}\max_j |f_j^k|\tau \\ &\le C, \end{aligned}$$

pričom konštanta C závisí len od ϕ a f.

Konvergencia metódy

Aproximácia sa overuje zvlášť pre $\dfrac{\partial}{\partial t}$ a pre $\dfrac{\partial^2}{\partial x^2}$. Postup je podobný ako pre diferenčnú schému pre obyčajnú diferenciálnu rovnicu 2. rádu. Ak riešenie $u(t,x)$ je dostatočne hladké, t.j. ak parciálne derivácie $\dfrac{\partial^2 u(t,x)}{\partial t^2}$ a $\dfrac{\partial^4 u(t,x)}{\partial x^4}$ sú ohraničené, tak sa dá dokázať

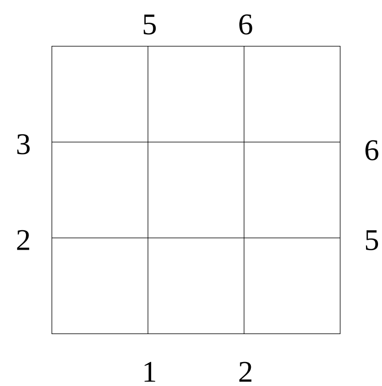
$$\max_{i,j} |u_j^i - u(t_i, x_j)| = \mathcal{O}\left(\tau + h^2\right).$$

Všimnime si, že časový a priestorový krok sú pre implicitnú schému navzájom nezávislé.

Cvičenie 10.1 Riešenie rovnice

$$-\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = 0$$

v $(0,3) \times (0,3)$ nadobúda dané hodnoty na stranách štvorcovej siete s krokom dĺžky $\Delta x = \Delta y = 1$ - pozri obrázok 10.4. Nájdite hodnoty riešenia vo vnútorných uzloch siete.

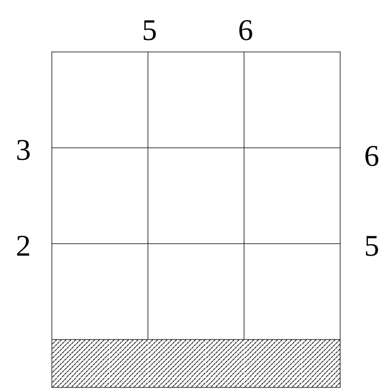


Obr. 10.4. Cvičenie pre Dirichletovu úlohu

Cvičenie 10.2 Riešenie rovnice

$$-\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = 0$$

v $(0,3) \times (0,3)$ nadobúda dané hodnoty na stranách štvorcovej siete s krokom dĺžky $\Delta x = \Delta y = 1$ - pozri obrázok 10.5. Šrafovanie znamená neumanovskú okrajovú podmienku $\nabla u \cdot \boldsymbol{\nu} = 0$, pričom $\boldsymbol{\nu}$ je vektor vonkajšej normály k danej strane. Nájdite hodnoty riešenia vo vnútorných uzloch siete.



Obr. 10.5. Cvičenie pre zmiešanú úlohu

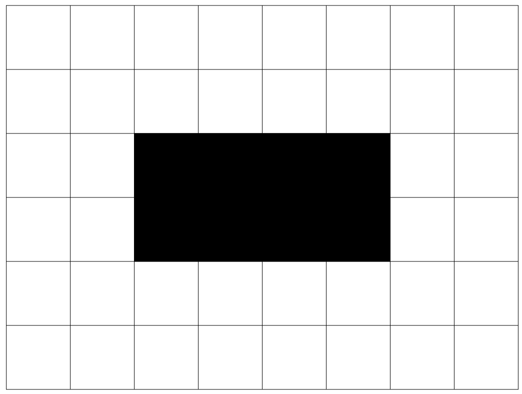
Cvičenie 10.3 Riešenie rovnice

$$-\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = 0$$

nadobúda konštantnú hodnotu $u = 100$ na vonkajších stranách štvorcovej siete s krokom $\Delta x = \Delta y = 1$ - pozri obrázok 10.6. V strede oblasti sa nachádza obdĺžnikový výrez znázornený čiernou farbou, na hranici ktorého platí okrajová podmienka tretieho druhu

$$-\nabla u(x, y) \cdot \boldsymbol{\nu} = u(x, y),$$

pričom $\boldsymbol{\nu}$ je vektor vonkajšej normály k danej strane. Napíšte diferenčnú schému pre približný výpočet riešenia na danej sieti.



Obr. 10.6. Cvičenie pre okrajovú podmienku tretieho druhu

Literatúra

1. ARSENIN, V. J.: *Matematická fyzika.* Bratislava, Alfa 1977.
2. ATKINSON, K. E.: *An Introduction to Numerical Analysis.* J. Wiley & Sons 1989.
3. BEREZIN, N. P., ŽIDKOV, I. S.: *Metody vyčislenij I,II.* Moskva, Fizmatgiz 1966.
4. COLLATZ, L.: *Funkcionální analýza a numerická matematika.* Praha, SNTL 1970.
5. FADDEJEV, V. N., FADDEJEVOVÁ, D. K.: *Numerické metody lineární algebry.* Praha, SNTL 1964.
6. GILBARG, N. S., TRUDINGER, D.: *Elliptic Partial Differential Equations of Second Order.* Berlin, Heidelberg, Springer 1983.
7. HERZBERGER, J.: *Übungsbuch zur Numerischen Mathematik.* Vieweg 1998.
8. ISAACSON, H. B., KELLER, E.: *Analyse numerischer Verfahren.* Leipzig, Verlag für Kunst und Wissenschaft 1972.
9. JOURNEL, A. G., HUIJBREGTS, Ch. J.: *Mining Geostatistics.* Academic Press 1978.
10. KARLIN, S., TAYLOR, H. M.: *A First Course in Stochastic Processes.* Academic Press 1975.
11. NEKVINDA, J., ŠRUBAŘ, M., VILD, J.: *Úvod do numerické matematiky.* Praha, SNTL 1976.
12. RALSTON, A.: *Základy numerické matematiky.* Praha, Academia 1973.
13. REKTORYS, K. a kolektív: *Přehled užité matematiky.* Praha, SNTL 1981.
14. SCHWARZ, H. R.: *Numerische Mathematik.* Stuttgart, B.G. Teubner 1993.
15. STAFFANS, O. J.: An inequality for positive definite Volterra kernels. *Proc. American Math. Society*, 58:205–210, 1976.
16. VITÁSEK, E.: *Numerické metody.* Praha, SNTL 1987.