Matematika 4

Ľ. Marko

November 14, 2019

CONTENTS

PREDHOVOR

Matematika je univerzálny jazyk pre fyzikálne a technické vedy. Preto je nutné aby študenti FEI STU Bratislava rozumeli základným pojmom a metódam numerickej matematiky. Predmet Matematika 4 pozostáva z dvoch častí: z úvodov do numerickej matematiky a matematickej štatistiky. Tento učebný text z časti numerická matematika som vytvoril po novej akreditácii pre študijný odbor JFI počas zimného semestra školského roku 2018/2019. Nepovažujem ho za konečnú verziu. Predmet "Matematika 4" sa bude dopĺňať a adaptovať v nasledujúcich rokoch. Ľ. Marko

Part I

Numerická matematika.

Základy numerickej matematiky môžu študentom uľahčiť riešenie mnohých praktických úloh. Tieto poznatky patria k "povinnej výbave" každého študenta FEI STU.

Chapter 1 DESATINNÉ MIESTA A PLATNÉ ČíSLICE.

Mnoho problémov, ktoré sa vyskytujú v inžinierkych oblastiach alebo vo fyzike nemá analytické riešenie a ak aj nejaké riešenie má, často je v takej forme, že sú ťažkosti ak vyžadujeme numerické výsledky. Sú mnohé dôvody prečo existujú také obmedzenia. Napríklad, nulové body funkcie obsiahnutej v riešení nie je možné vypočítať analyticky, určitý integrál nemožno nájsť analyticky, alebo nemožno nájsť analytické riešenie nelineárnej diferenciálnej rovnice, alebo je potrebné poznať riešenie veľkých sústav lineárnych rovníc.

Situácia iného typu nastáva, ak je známe analytické riešenie, ale jeho aplikácia v špeciálnom prípade vedie na "zakázané" množstvo výpočtov, teda je nutné nájsť efektívnejšie a účinnejšie numerické metódy.

Pretože väčšina numerických výsledkov iba aproximuje výpočty, v ktorých sa používajú $\sqrt{2}, e$ alebo π je nutné mať jednoduché pravidlá ako určiť ich presnosť. Toto možno dosiahnuť konštatovaním, že výsledok je presný na n desatinných miest, alebo že je presný na daný počet dôležitých číslic. Napríklad počet desatinných miest ak aproximujeme napríklad číslo $17,213622$, na tri desatinné miesta, skúmame štvrtú číslicu za desatinnou čiarkou a ak je to číslo 5 alebo viac predchádzajúcu číslicu zväčšíme o jednu a výsledok odstrihneme (skrátime) na tri desatinné čísla. Avšak ak je štvrtá číslica 4 alebo menej, predchádzajúce číslo sa ponechá bez zmeny a výsledok skrátime na existujúce tri čísla po desatinnej čiarke. Ak tento proces aplikujeme na predchádzajúce číslo a aproximujeme ho s presnosťou na tri desatinné miesta potom to bude $17,214$ zatiaľ čo ak ho aproximujeme s presnosťou na štyri desatinné miesta, potom to bude $17,2136$. Tento proces aproximácie čísel na n desatinných miest rastom $n-$ tého čísla o 1, ak $(n+1)-$ vá číslica je 5 alebo viac a potom odstrihnutím výsledku po n desatinných miestach sa nazýva zaokrúhľovaním výsledku na n desatinných miest nahor. Podobne proces ponechania $n-$ tého čísla nezmeneného ak $(n+1)-$ vá číslica je 4 alebo menej zaokrúhľovaním nadol na n desatinných miest.

Vyjadriť číslo s presnosťou na n platných číslic vyžaduje trochu iné argumenty ako vyššie. Prvá nenulová číslica vyskytujúca sa v čísle nezávisle od polohy desatinnej čiarky sa nazýva prvá platná (signifikantná) číslica, teda napríklad pre číslo $3,496221$ je takou číslo 3 a pre číslo $0,004713$ je to 4 . Ak počítame od prvej platnej číslice $(n+1)$ číslic doprava, $n-$ tá číslica sa zaokrúhľuje nahor, alebo nadol podľa $(n+1)-$ vej číslice. Číslo odstrihnuté za skupinou n čísel nahradením každého čísla pred desatinnou čiarkou nulou je potom nazývame vyjadrením čísla s presnosťou na n platných čísel. Tak číslo $315,814$ bude na tri platné čísla zmenené na $316,000$ zatiaľ čo číslo $0,004723217$ na štyri platné čísla bude $0,004723$.

Presnosť sa môže stratiť, ak použijeme približný výsledok numerických výpočtov, alebo pri predchádzajúcich numerických výpočtoch sa proces opakuje mnoho krát. Aby sme predišli strate presnosti je nutné pracovať s fixným počtom číslic, ktoré je dostatočne veľké. Kalkulačky a počítače používajú pevný počet číslic výpočtové balíky symbolických algebier dávajú možnosť výberu počtu, aby sa dosiahla vysoká presnosť pri postupnosti výpočtov.

Tvar, v ktorom reprezentujeme čísla doteraz sa nazýva pevná rádová čiarka. Plávajúca desatinná čiarka (fixed and floating point numbers decimal representation) používaná vo väčšine počítačov dovoľuje zapísať číslo x v tvare $x = r.N^s$, kde číslo N sa nazýva základ reprezentácie, číslo r sa nazýva mantisa a s exponent. Mantisa sa zvyčajne vyberá, aby mala jedno číslo pred desatinnou čiarkou. Teda ak je základ 10, číslo $453,7$ má v plávajúcej desatinnej čiarke reprezentáciu $4,537 \times 10^2$, a číslo $0,000369$ má reprezentáciu $3,69 \times 10^{-4}$. Označenie používané pre reprezentáciu v plávajúcej desatinnej čiarke v strojových výpočtoch so základom 10 dovoľuje reprezentovať x vo forme s plávajúcou desatinnou čiarkou, zápisom mantisy a potom symbol E nasledovaným kladným alebo záporným exponentom. Väčšina počítačov je normalizovaná tak, že mantisa je medzi 0 a 1, teda ak použijeme túto konvenciu číslo $453,7$ bude v tvare $0,4537E3$, a $0,000369$ bude $0,369E-3$.

Chapter 2 KORENE NELINEÁRNYCH FUNKCIí.

Nech $f(x)$ je spojitá, reálna funkcia definovaná na $a \leq x \leq b$. Číslo ξ sa nazýva koreň funkcie $f(x)$ v tomto intervale $\langle a, b\rangle$ ak $f(\xi) = 0$ a číslo $x = \xi$ sa nazýva nulový bod $f(x)$. Hľadanie koreňov funkcií je základom rozvoja a aplikácií matematiky a iba v jednoduchých prípadoch možno nájsť korene analyticky, teda v ostatných prípadoch sa hľadajú numericky. Existuje mnoho numerických metód hľadania koreňov, ale my sa budeme zaoberať metódou bisekcie, metódou pevného bodu a Newtonovou metódou, pretože sú pomerne jednoduché a ľahko ich možno implementonať na počítače. Riešiť numericky rovnicu $f(x) = 0$ znamená aproximovať jej koreň ξ s vopred danou toleranciou. Metódy jej riešení z hľadiska podmienok a rýchlosti konvergencie charakterizujeme ako

- štartovacie (nenáročné podmienky konvergencie, ale pomalá konvergencia),
- spresňujúce (zložitejšie podmienky konvergencie, ale rýchla konvergencia).

Numerický výpočet potom môže prebiehať tak, že niekoľko krokov sa vykoná štartovacou metódou a potom sa pokračuje spresňujúcou metódou.

Štartovacie metódy:

Grafická metóda.

Presné nakreslenie grafu funkcie $y = f(x)$ môže pomôcť lokalizovať reálne korene rovnice $f(x) = 0$ t.j. určiť intervaly, v ktorých korene určite ležia. Pretože korene rovnice $f(x) = 0$ sú $x-$ ové súradnice priesečníku grafu funkcie $y = f(x)$ s osou o_x môžeme si urobiť dosť konkrétnu predstavu o ich polohe. Niekedy je výhodnejšie rovnicu $f(x) = 0$ písať v tvare $f_1(x) = f_2(x)$. Grafická metóda nám tiež môže poskytnúť informáciu, či reálny koreň danej rovnice v uvažovanom untervale skutočne existuje. Pre zložitejšie typy funkcií môžme lokalizovať korene tabelovaním funkcie $y = f(x)$.

Metóda bisekcie.

Okrem grafu $f(x)$ a hľadania tých hodnôt x, pre ktoré je $f(x) = 0$, je to najjenoduchšia systematická metóda hľadania koreňov funkcie $f(x)$. Metóda sa dá ľahko naprogramovať a aplikuje sa pri hľadaní koreňov rovníc funkcie $f(x)$, ktorá má vlastnosť zmeny znamienka keď x prechádza cez koreň. Určenie koreňa touto metódou presne závisí od možnosti vypočítať funkciu s dostatočnou presnosťou t.j. zmena znamienka funkcie musí byť určená korektne. Aby sme pochopili ako metóda pracuje uvažujme spojitú funkciu $f(x)$ a čísla $\alpha < \beta$, také, že $f(\alpha)$ a $f(\beta)$ majú opačné znamienka. Podľa vety o medzihodnote musí funkcia $f(x)$ mať aspoň jeden koreň ξ medzi α a β, ako na obrázku. Avšak ako je vidieť na ďalších obrázkoch ak $f(\alpha)$ a $f(\beta)$ majú rovnaké znamienka, nič nemožno povedať o existencii koreňov v danom intervale, či existuje dvojnásobný koreň, dva korene, alebo

žiadny koreň neexistuje. Budeme teda predpokladať, že $f(x)$ prechádza zmenou znamienka krížom cez interval a že α a β sú vybrané dostatočne blízko tak, že vo vybranom intervale je iba jeden koreň. Keď je $f(x)$ dostatočne jednoduchá dá sa to dosiahnuť grafom $f(x)$ a výberom vhodných hodnôt pre α a β. Pre implementáciu metódy bisekcie je najjednoduchšie vidieť kde funkcia $f(x)$ má opačné znamienka na intervale $\alpha < x < \beta$. Také testy sa dajú urobiť skúmaním znamienka súčinu $f(\alpha)f(\beta)$. Ak nastane situácia, keď počas výpočtu metódou bisekcie počítač vypočíta $f(\alpha)f(\beta) = 0$ treba overiť, hodnotu $f(\alpha)$ aby sme sa vyhli interpretácii α ako skutočnej nuly, kedže približnú hodnotu α môže interpretovať ako nulu. Prvý krok v metóde bisekcie zvolíme ako rozpolenie (bisekciu) intervalu $\alpha \leq x \leq \beta$ na dva podintervaly $\alpha < x < x_1$ a $x_1 < x < \beta$, kde $x_1 = \frac{1}{2}(\alpha + \beta)$. Podinterval, ktorý budeme brať do úvahy v ďalšom kroku dostaneme zámenou α namiesto x_1 ak $f(\alpha)f(x_1) > 0$, pretože v takom prípade $f(x)$ mení znamienko v podintervale $x_1 < x < \beta$ teda tento interval musí obsahovať koreň $f(x)$. Opačne ak $f(\alpha)f(x_1) < 0$, budeme uvažovať podinterval, v ktorom zameníme β za x_1, pretože v tomto prípade $f(x)$ prechádza zmenou znamienka v intervale $\alpha < x < x_1$, a tak tento interval musí obsahovať koreň ξ. Tak sme úlohu hľadania koreňa na intervale $\alpha \leq x \leq \beta$ zamenili za úlohu hľadania koreňa na "polovičnom" intervale. Metóda bisekcie spočíva v opakovaní tejto procedúry vždy na menšom intervale tak, že po m krokoch koreň ξ bude v intervale dĺžky $\frac{|\alpha-\beta|}{2^m}$. Ak požadujeme, aby koreň bol nájdený s presnosťou (chybou, toleranciou) ε, kde $\varepsilon > 0$ je dopredu určená malá hodnota, počítačové výpočty ktoré pracujú s pevným počtom číslic budú pracovať pokiaľ postupné iterácie x_m a x_{m+1} nesplnia podmienku $|x_m - x_{m+1}| < \varepsilon$. Požadovaná aproximácia koreňa ξ je potom $x_m \pm \varepsilon$. Metóda bisekcie má vlastnosť, že chyba sa zmenšuje na polovicu po každej iterácii. Na rozdiel od iných metód ak použijeme metódu bisekcie tá vždy konverguje ku koreňu, aj keď v počiatočnom intervale $\alpha \leq x \leq \beta$ je viac koreňov dopredu nikdy nevieme, ku ktorému koreňu bude konvergovať.

Podmienky úlohy.

- $\langle \alpha, \beta \rangle \subset \boldsymbol{R}$ je uzavretý interval,

- rovnica je v tvare $f(x) = 0$, kde f je spojitá funkcia na intervale $\langle \alpha, \beta \rangle$,

- f má na intervale $\langle \alpha, \beta \rangle$ práve jeden koreň ξ, t.j. existuje práve jedno číslo $\xi \in \boldsymbol{R}$, pre ktoré $f(\xi) = 0$,

- funkčné hodnoty majú v koncových bodoch intervalu opačné znamienka, t.j. $f(\alpha)f(\beta) < 0$,

- ε je prípustná tolerancia chyby numerického riešenia ξ.

Algoritmus metódy bisekcie:

Vstup: $f, \langle x_l = \alpha, x_r = \beta \rangle, \varepsilon$,

1. $x_n = \frac{x_l + x_r}{2}$,
2. ak $f(x_n) = 0, \xi := x_n$, koniec,
3. ak $x_n - \alpha < \varepsilon$, $\xi := x_n$, koniec,
4. ak $f(x_n)f(\alpha) < 0, \beta := x_n$, choď na 1.
5. inak $\alpha := x_1$, choď na 1.

Výstup ξ.

Výhodou metódy bisekcie je jej jednoduchosť a to, že používa minimálne množstvo informácií, pretože závisí iba od hodnôt funkcie $f(x)$ v koncových bodoch intervalu a nie od derivácií, aj keď iné metódy môžu konvergovať rýchlejšie. Praktická im-

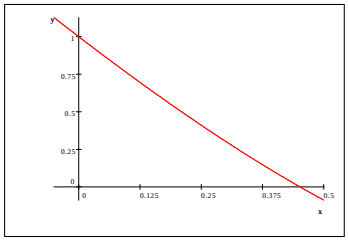
plementácia mb. na počítači môže trpieť faktom, že ak vypočítaný súčin $f(\alpha)f(\beta)$ podtečie pohyblivú desatinnú čiarku budú nevyhnutne hornou a dolnou hranicou koreňa. Avšak toto je možné prekonať určovaním znamienka $f(\alpha)f(\beta)$ tak, že sa určia znamienka $f(\alpha)$ a $f(\beta)$. Pretože mb. je málo ovplyvnená limitnou presnosťou na začiatku výpočtu sa často používa iná a rýchlejšia metóda a prepína sa na mb. ak chceme dostať veľmi presnú aproximáciu koreňa mb. nemožno použiť na hľadanie koreňa $x = \xi$ funkcie, ktorá je buď konvexná alebo konkávna v bode $x = \xi$, alebo nemení znamienko keď x prechádza cez ξ. To sa môže stať ak hľadáme korene polynómu párneho stupňa najjednoduchším je napríklad $f(x) = (x-a)^2$ s dvojnásobným koreňom a.

Numerické určenie násobného koreňa je ťažké, preto iba načrtneme možný prístup pre polynóm stupňa n s reálnymi koreňmi, z ktorých je jeden dvojnásobný. Problémy sa vyskytujú keď hľadáme násobné korene, pretože výpočty vždy vedú na zle podmienený problém - t.j. keď extrémne malá chyba vo výpočte vedie ku extrémne veľkej chybe vo výsledku.

Prístup, ktorý popíšeme sa nazýva deflácia polynómu. Najprv nájdeme jednoduchý koreň polynómu, a polynóm potom vydelíme odpovedajúcim faktorom, aby sme dostali polynóm stupňa $n-1$. Opakovaním tohto procesu nájdeme všetkých $n-2$ jednoduchých koreňov polynómu, čo vedie ku kvadratickému polynómu s dvojnásobným koreňom, ktorý možno nájsť pomocou formuly pre riešenie kvadratickej rovnice. Ak je to nutné treba urobiť defláciu, aby sa zamedzilo compounding of chýb. Je dôležité mať na zreteli, že mb. sa nedá použiť na hľadanie koreňov párnych rádov, pretože v týchto prípadoch sa nedeje zmena znamienka, ale pracuje dobre pre korene nepárneho rádu nezávisle od ich rádu. Možný spôsob ich prekonania pri hľadaní násobných koreňov je použitie mb. s rôznymi štartovacími intervalmi, ďalším je použitie iných metód s rôznymi odhadmi.

Example 1 *Použitím metódy bisekcie nájdime najmenší koreň funkcie* $f(x) = 1 - 3x + \frac{1}{2}xe^x$.

Solution 2



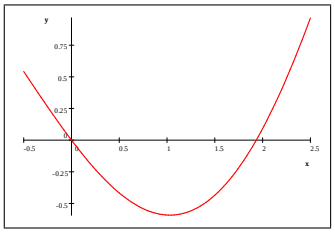
Obrázok grafu funkcie ukazuje, že aproximácia najmenšieho koreňa $f(x) = 0$ *je* $x = 0,45$ *a vhodné hodnoty pre* α *a* β *sú* $\alpha = 0,43$ *a* $\beta = 0,47$. *Potom* $f(\alpha) = 0,0405$ *a* $f(\beta) = -0,0340$ *a graf ukazuje, že medzi* α *a* β *existuje iba jeden koreň. Ak každý krok výpočtu ľavý koncový bod intervalu obsahujúci koreň* ξ *označíme* x_l *a pravý koncový bod* x_r *výpočet môže prebiehať nasledovne:*

n	x_l	x_r	x_n	$f(x_l)$	$f(x_n)$	$f(x_l)\,f(x_n)$
1	$\alpha = 0,43$	$\beta = 0,47$	$0,45$	$0,0405$	$0,0029$	> 0
2	$0,45$	$0,47$	$0,46$	$0,0029$	$-0,0157$	< 0
3	$0,45$	$0,46$	$0,455$	$0,0029$	$-0,0064$	< 0
4	$0,45$	$0,455$	$0,4525$	$0,0029$	$-0,0018$	< 0

n	*nový interval*	*aproximácia koreňa*
1	$0,45 < \xi < 0,47$	$0,45$
2	$0,45 < \xi < 0,46$	$0,46$
3	$0,45 < \xi < 0,455$	$0,455$
4	$0,45 < \xi < 0,4525$	$0,4525$

Pokračovanie v tomto procese ukáže, že ak $\varepsilon = 0,00001$ *s presnosťou na päť desatinných miest požadovaná hodnota koreňa bude* $x = 0,45154$.

Example 3 *Použitím metódy bisekcie nájdime najmenší kladný koreň funkcie* $f(x) = \left(\frac{x}{2}\right)^2 - \sin x$.

Solution 4 

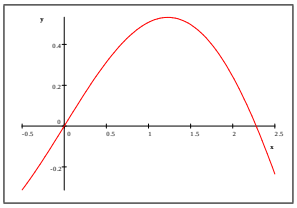
Obrázok grafu funkcie ukazuje, že aproximácia najmenšieho kladného koreňa $f(x) = 0$ *je* $x = 1.945$ *a vhodné hodnoty pre* α *a* β *sú* $\alpha = 0,43$ *a* $\beta = 0,47$. *Potom* $f(\alpha) = 0,0405$ *a* $f(\beta) = -0,0340$ *a graf ukazuje, že medzi* α *a* β *existuje iba jeden koreň. Ak každý krok výpočtu ľavý koncový bod intervalu obsahujúci koreň* ξ *označíme* x_l *a pravý koncový bod* x_r *výpočet môže prebiehať nasledovne:*

n	x_l	x_r	x_n	$f(x_l)$	$f(x_n)$	$f(x_l)\, f(x_n)$
1	$\alpha = 1,8$	$\beta = 2$	$1,9$	$-0.163\,85$	$-0.043\,8$	> 0
2	1.9	2	1.95	$-0.043\,8$	$2.166\,5 \times 10^{-2}$	< 0
3	1.9	1.95	1.925	$-0.043\,8$	$-1.151\,7 \times 10^{-2}$	> 0
4	1.925	1.95	1.9375	$-1.151\,7 \times 10^{-2}$	$4.962\,3 \times 10^{-3}$	< 0
5	1.925	$1,9375$	$1.931\,3$	$-1.151\,7 \times 10^{-2}$	$-3.239\,4 \times 10^{-3}$	> 0
6	$1.931\,3$	$1,9375$	$1.934\,4$	$-3.239\,4 \times 10^{-3}$	$8.545\,7 \times 10^{-4}$	< 0

n	*nový interval*	*aproximácia koreňa*
1	$1.9 < \xi < 2$	1.9
2	$1,9 < \xi < 1,95$	$1,95$
3	$1.925 < \xi < 1.95$	1.925
4	$1,925 < \xi < 1,9375$	1.9375
5	$1.931\,3 < \xi < 1,9375$	$1.931\,3$
6	$1.934\,4 < \xi < 1,9375$	$1.934\,4$

Pokračovanie v tomto procese ukáže, že ak $\varepsilon = 0,00001$ *s presnosťou na päť desatinných miest požadovaná hodnota koreňa bude* $x = 0,45154$.

Example 5 *Použitím metódy bisekcie nájdime najmenší kladný koreň funkcie* $f(x) = \sin x - \frac{1}{3}x = 0$ *blízko* $x = 2.2$. $[2.27886]$.

Solution 6 

Obrázok grafu funkcie ukazuje, že aproximácia najmenšieho kladného koreňa $f(x) = 0$ *je* $x = 2.2$ *a vhodné hodnoty pre* α *a* β *sú* $\alpha = 2,1$ *a* $\beta = 2,4$. *Potom* $f(\alpha) = 0,0405$ *a* $f(\beta) = -0,0340$ *a graf ukazuje, že medzi* α *a* β *existuje iba jeden koreň. Ak každý krok výpočtu ľavý koncový bod intervalu obsahujúci koreň* ξ *označíme* x_l *a pravý koncový bod* x_r *výpočet môže prebiehať nasledovne:*

n	x_l	x_r	x_n	$f(x_l)$	$f(x_n)$
1	2.278 2	2.280 6	2.279 4	$6.516\,9 \times 10^{-4}$	$-5.286\,9 \times 10^{-4}$
2	2.278 2	2.279 4	2.278 8	$6.516\,9 \times 10^{-4}$	$6.163\,7 \times 10^{-5}$
3	2.278 8	2.279 4	2.279 1	$2.807\,3 \times 10^{-2}$	$-2.334\,9 \times 10^{-4}$
4	2.25	2.287 5	2.268 8	$2.807\,3 \times 10^{-2}$	$9.860\,1 \times 10^{-3}$
5	2.268 8	2.287 5	2.278 2	$9.860\,1 \times 10^{-3}$	$6.516\,9 \times 10^{-4}$
6	2.278 2	2.287 5	2.282 9	$6.516\,9 \times 10^{-4}$	$-3.977\,7 \times 10^{-3}$
7	2.278 2	2.282 9	2.280 6	$6.516\,9 \times 10^{-4}$	$-1.710\,2 \times 10^{-3}$

n	$f(x_l)\, f(x_n)$	*nový interval*
1	$<$	2.278 2, 2.279 4
2	$>$	2.278 8, 2.279 4
3	$<$	2.25, 2.287 5
4	$>$	2.268 8, 2.287 5
5	$>$	2.278 2, 2.287 5
6	$<$	2.278 2, 2.282 9
7	$<$	2.278 2, 2.280 6

Pokračovanie v tomto procese ukáže, že ak $\varepsilon = 0,00001$ *s presnosťou na päť desatinných miest požadovaná hodnota koreňa bude* $x = 0,45154$.

Metóda prostej iterácie.

Metóda prostej iterácie. Táto metóda je vhodná pre počítačové výpočty za predpokladu, že numerické hodnoty funkcií, ktoré počítame sa dajú jednoducho vypočítať a dobrá aproximácia koreňa sa použije ako štart iteračného procesu. Rovnicu $f(x) = 0$ prepíšeme na tvar (býva väčšinou niekoľko možností) $x = g(x)$ a predpokladáme, že existuje interval $I_0 = \langle a_0, b_0 \rangle$ patriaci k definičnému oboru a oboru spojitosti ako funkcie f tak aj funkcie g taký, ktorý obsahuje spoločný koreň ξ oboch uvedených rovníc. Myšlienka je jednoduchá a jej úspech závisí od prepísania danej funkcie $f(x)$, ktorej koreň zisťujeme, na tvar $f(x) = x - g(x)$. Potom ak $x = \xi$ anuluje výraz $x - g(x)$, tak ξ je koreň $f(x)$. Reprezentácia $f(x)$ na tvar $f(x) = x - g(x)$ nie je jediná, pretože ako vidieť na nasledujúcom príklade $g(x)$ možno napísať viac ako jedným spôsobom. Neskôr odvodíme jednoduchú podmienku pre funkciu $g(x)$, ktorá musí byť splnená spolu s hodnotou $x_0 = \alpha$ použitou na štart iteračného procesu, aby výpočty konvergovali ku koreňu ξ. Ak teraz uvažujeme funkciu $g(x)$ ako zobrazuje bod x na bod $g(x)$, potom koreň $x = \xi$ rovnice $f(x) = x - g(x) = 0$ má vlastnosť, že $g(x)$ zobrazí bod ξ na seba a to je dôvod, prečo ξ nazývame pevný bod rovnice

$$x = g(x) \qquad ((1))$$

Pevný bod iteračnej schémy vyplývajúej z (1) ak ju zapíšeme ako pevný bod a iterácia

$$x_{n+1} = g(x_n) \qquad ((2))$$

a štartujeme iteračný proces krokom $x_0 = \alpha$. Iterácia konverguje ak postupnosť iterácií x_n má limitu ak $n \to \infty$, a diverguje ak taká limita neexistuje. Predpokladáme, že ak iterácie konvergujú k výsledku s požadovanou presnosťou (chybou) ε, kde $\varepsilon > 0$ je predpísané malé číslo a podmienka $|x_m - x_{m+1}| < \varepsilon$. Požadovaná aproximácia koreňa ξ je potom $x_n \pm \varepsilon$.

Example 7 *Nájdime pevný bod iteračnej schémy na výpočet* $\sqrt{a}$, *keď* $a > 0$, *a použite ju na výpočet* $\sqrt{2}$ *s presnosťou na 6 desatinných miest.*

Solution 8 *Požadované číslo* $\sqrt{a}$ *je riešenie rovnice* $x^2 = a$, *teda vyjadriť ho v tvare (1) prepíšeme ju ako* $2x^2 = x^2 + a$, *a potom delíme výsledok* $2x$ *a dostaneme* $x = \frac{1}{2}\left(x + \frac{a}{x}\right)$, *teda v označení (1) funkcia* $g(x) = \frac{1}{2}(x + \frac{a}{x})$. *Pevný bod iteračnej schémy vyplýva z (1), zámenou* x *na ľavej strane na* x_{n+1} *a* x *na pravej strane na* x_n *a dostaneme* $x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$. *Iteráciu štartujme pre* $n = 0$ *a* $x_0 = k$, *kde* k *je aproximácia* $\sqrt{a}$. *Aby sme ilustrovali schému výpočtu* $\sqrt{2}$, *teda schéma bude* $x_{n+1} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right)$, *a pre jednoduchosť štartujeme s* $x_0 = 1$. *Výsledky kalkulácií sú*

$x_0 = 1,$
$x_1 = 1.5,$
$x_2 = 1.41666667,$
$x_3 = 1.41421569,$
$x_4 = 1.41421356,$
$x_5 = 1.41421356.$

Pretože x_4 *a* x_5 *sú identické zaokrúhlime výsledok* x_5 *na šesť desatinných miest a teda* $\sqrt{2} = 1.414214$. *Prostá iteračná schéma v tomto príklade konverguje rýchlo a je to táto metóda , ktorá sa používa v počítačoch na určenie druhej odmocniny z ľubovoľného kladného reálneho čísla s presnosťou vo vnútri výpočtového systému a použitého softwéru. Experimentovanie ukazuje, že iteračná schéma je stabilná bez ohľadu na výber štartovacieho bodu, pretože bude vždy konvergovať ku* $\sqrt{2}$, *ale keď štartovací bod bude blízko* $\sqrt{2}$, *potom je konvergencia bude najrýchlejšia.*

V nasledujúcom príklade preskúmame iteračnú schému trochu viac a ukážeme, že konvergencia sa nie vždy podarí.

Example 9 *Navrhnime iteračnú schému a nájdime korene kvadratickej rovnice* $2x^2 - 24x + 41 = 0$ *a testujme ich numericky.*

Solution 10 *Je možné navrhnúť dve iteračné schémy priamo z rovnice:* $x = \frac{1}{24}(2x^2 + 41)$, *alebo* $x = 12 - \frac{41}{2x}$, *odkiaľ dostaneme:*

Schéma A: $x_{n+1} = \frac{1}{24}\left(2x_n^2 + 41\right)$, *a*

Schéma B: $x_{n+1} = 12 - \frac{41}{2x_n}$. *Použitím riešenia kvadratickej rovnice dostaneme korene :* $x = 6 - \frac{1}{2}\sqrt{62} = 2.0630$ *a* $x = 6 + \frac{1}{2}\sqrt{62} = 9.9370$, *teda štartovacie aproximácie blízko týchto hodnôt sú*

Schéma A		*Schéma B*	
$x_0 = 2$	$x_0 = 10$	$x_0 = 2$	$x_0 = 10$
$x_1 = 2.0417$	$x_1 = 10.0417$	$x_1 = 1.75$	$x_1 = 9.7222$
$x_2 = 2.0557$	$x_2 = 10.1113$	$x_2 = 0.2857$	$x_2 = 9.8914$
$x_3 = 2.0605$		$x_3 = -59.7500$	$x_3 = 9.9275$
$x_4 = 2.0621$		$x_4 = 12.3431$	$x_4 = 9.9350$
$x_5 = 2.0627$		$x_5 = 10.3392$	$x_5 = 9.9370$
$x_6 = 2.0630$		$x_6 = 10.0172$	$x_6 = 9.9370$
...		$x_7 = 9.9535$	
	$x_8 = 12.4801$	...	...
	$x_9 = 14.6877$		
$x_\infty = 2.0630$	$x_\infty = \infty$	$x_\infty = 9.9370$	$x_\infty = 9.9370$

Teda Schéma A je iba čiastočne úspešná, pretože aj keď štartuje s $x_0 = 2$ *konverguje k nule blízko* 2, *a diverguje keď štartuje s* $x_0 = 10$. *Podobne aj Schéma B je čiastočne úspešná aj keď z iného dôvodu. Aj keď iterácie konvergujú ku koreňu blízko* 10, *ak štartujú s* $x_0 = 9$, *ak štartujú s* $x_0 = 2$ *nekonvergujú ku koreňu blízko* 2 *ale opäť ku koreňu blízko* 10. *Aby sme pochopili chovanie iteračných schém, nasledujúca veta nám ukáže podmienky, na výber* $g(x)$ *a štartovacej aproximácie* x_0, *ktoré zaručia konvergenciu iteračnej schémy.*

Theorem 11 *Konvergencia iteračnej schémy. Nech* $g(x)$ *je definovaná na intervale* $a \leq x \leq b$, *v ktorom má pevný bod* ξ, *a nech* $g(x)$ *je spojitá na tomto intervale so spojitou deriváciou* $g'(x)$ *takou, že* $|g'(x)| \leq k < 1$. *Potom rovnica* $x = g(x)$ *má jediný pevný bod* ξ *v intervale* a *ak* x_0 *je také, že* $a \leq x_0 \leq b$ *iteračná schéma* $x_{n+1} = g(x_n)$ *konverguje ku* ξ.

Dôkaz Nech by existovali dva rôzne pevné body ξ_1 a ξ_2 v intervale, teda $\xi_1 = g(\xi_1)$ a $\xi_2 = g(\xi_2)$. Potom aplikáciou vety o strednej hodnote a podmienky $|g'(x)| \leq k < 1$, dostaneme $|\xi_1 - \xi_2| = |g(\xi_1) - g(\xi_2)| = |g'(\eta)(\xi_1 - \xi_2)| \leq k\,|\xi_1 - \xi_2| < |\xi_1 - \xi_2|$, čo je spor. Na dôkaz konvergencie použijeme opäť vetu o strednej hodnote: existuje ζ_n medzi x_{n-1} a ξ taký, že $|\xi - x_n| = |g(\xi) - g(x_{n-1})| = |g'(\zeta_n)(\xi - x_{n-1})| = |g'(\zeta_n)||\xi - x_{n-1}| \leq x_0|\xi - x_{n-1}|$. Opakovaná aplikácia tejto nerovnice vedie na výsledok $|\xi - x_n| \leq x_0^n|\xi - x_0|$, ale pretože $0 \leq x_0 < 1$ máme $\lim_{n\to\infty} x_0^n = 0$, teda $\lim_{n\to\infty}|\xi - x_n| = 0$, a teda $\lim_{n\to\infty} x_n = \xi$. S menšími ťažkosťami možno ukázať, že iterácie tvoria Cauchyho postupnosť a s ohľadom na úplnosť reálnych čísel máme, že postupnosť má limitu ξ, a veta je dokázaná.

Táto veta vysvetľuje výsledky príkladu. V Schéme A funkcia $g(x) = \frac{1}{24}(2x^2 + 41)$, teda $|g'(x)| = \frac{1}{6}|x|$ a $|g'(x)| < 1$ ak $0 < x < 6$, ukazuje, že schéma je konvergentná v blízkosti 2 ak sa použije začiatočná aproximácia blízko 2. Avšak ak $x = 10$ podmienka vety nie je splnená, teda schéma nemusí byť konvergentná ku koreňu blízko 10, aj keď nemôžme tvrdiť, že nebude konvergovať. V prípade Schémy B máme $g(x) = 12 - \frac{41}{2x}$, teda $|g'(x)| = \frac{41}{2x^2}$. Toto kazuje, že schéma bude konvergovať ku koreňu blízko 10 pre x_0 blízko 10, pretože potom $|g'(x)| < 1$, ale nemôžeme očakávať, že bude konvergovať ku koreňu blízko ku 2, kde je podmienka porušená, aj keď nemôžme tvrdiť, že nebude konvergovať. Je možné ukázať, že

ak $|g'(x)| > 1$, iterácia nebude konvergovať, iba ak náhodou. Dôvod konvergencie, alebo divergencie iteračnej schémy je najľahšie pochopiteľný použitím grafickej reprezentácie iteračného procesu.

Typickou spresňujúcou metódou je metóda dotyčníc.

Newtonova metóda.

Našou štartovacou metódou pre odvodenie Newtonovej metódy možno napísť v tvare

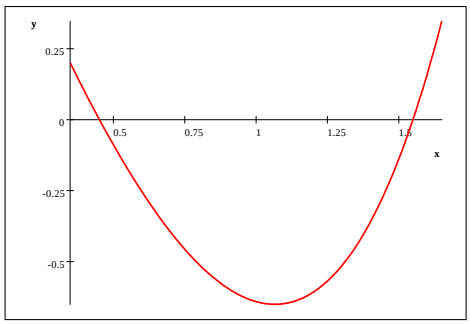
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(\xi)(x - x_0)^2, \qquad ((3))$$

kde ξ je bod medzi x_0 a x.

Linearizujeme $f(x)$ a položíme $f(x) = 0$, teda $f(x_0) + f'(x_0)(x_1 - x_0) = 0 \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$, potom nahradíme rovnicu $f(x) = 0$, rovnicou $f(x_1) + f'(x_1)(x - x_1) = 0$, koreňom tejto rovnice bude $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$. Iteráciou tohto výsledku dostaneme $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ dostaneme Newtonovu (Raphsonovu) metódu pre $n = 0, 1, 2, 3, \ldots$. Ak je daná tolerancia ε, kde $\varepsilon > 0$ je dané malé číslo, výpočty prebiehajú dovtedy, pokiaľ je prvý krát splnená podmienka $|x_m - x_{m+1}| < \varepsilon$. Číslo $x_{m+1} \pm \varepsilon$ považujeme za aproximáciu koreňa ξ. Poznamenajme, že Newtonova metóda je špeciálny prípad iteračnej metódy s $g(x) = x - \frac{f(x)}{f'(x)}$ a podľa predošlej vety máme $|\xi - x_n| = |g'(\zeta_n)||\xi - x_{n-1}|$, čo nám hovorí, že $|\xi - x_n|$ aproximuje $|g'(\zeta_n)||\xi - x_{n-1}|$ ak iterácia konverguje. Teda, čím menšie $|g'(\zeta_n)|$, tým rýchlejšia konvergencia. Pre Newtonovu metódu a jednoduchý koreň, je tento výraz nulový. Čo naznačuje, že pre Newtonovu metódu iteracie konvergujú rýchlejšie ako lineárne. Poznamenajme, že iteračná a Newtonova metóda konvergujú ku koreňu najbližšiemu k začiatočnému odhadu, čo nie je pravda pre m.b. Newtonova metóda je všeobecne rýchlejšie konvergujúca než m.b. pre jednoduché korene ale nie pre násobné korene.

Example 12 *Použite Newtonovu metódu na nájdenie núl* $f(x) = 1 - 3x + \frac{1}{2}xe^x$ *s presnosťou na päť desatinných miest.*

Solution 13 *Graf* $f(x) = 1 - 3x + \frac{1}{2}xe^x$



ukazuje, že má nuly blízko 0.5 *a* 1.6, *teda použijeme tieto ako naše štartovacie aproximácie. Pretože*

$f'(x) = \frac{1}{2}(1+x)e^x - 3$, Newtonova metóda bude $x_{n+1} = x_n - \frac{1-3x_n+\frac{1}{2}x_n e^{x_n}}{\frac{1}{2}(1+x_n)e^{x_n}-3}$ pre $n = 0, 1, 2, ...$ Štartujme výpočty s $x_0 = 0.5$ čo dáva

$x_0 = 0.5$

$x_1 = 0.450200$

$x_2 = 0.451541$

$x_3 = 0.451542$

$x_4 = 0.451542$, teda s presnosťou na 5 desatinných miest najmenší nulový bod $f(x)$ je 0.45154.Podobne ak výpočty štartujeme s $x_0 = 1.6$, dostaneme

$x_0 = 1.6$

$x_1 = 1.552769$

$x_2 = 1.549552$

$x_3 = 1.549538$

$x_4 = 1.549538$, teda s presnosťou na 5 desatinných miest najväčší nulový bod $f(x)$ je 1.54954.□

Tento príklad ilustruje rýchlosť s akou Newtonova metóda môže konvergovať k nule ak sa použije dobrá štartovacia aproximácia a dotyčnica ku grafu $y = f(x)$ v nulovom bode sa nenakláňa k dátam s malým uhlom k osi o_x, čo robí vyššiu presnosť nemožnou.

Cvičenia.

V cvičeniach 1 - 6 použite (metódu bisekcie) m.b. na nájdenie požadovaných koreňov

1. Koreň $\sin x - \frac{1}{3}x = 0$ blízko $x = 2.2$. [2.27886]
2. Koreň $e^{\frac{x}{3}} - x^2 = 0$ blízko $x = 1.1$.
3. Koreň $3\ln x + x^2 - 3 = 0$ blízko $x = 1.3$. [1.40619]
4. Najväčší kladný koreň $x^3 - 1.9x^2 - 2.3x + 3.7 = 0$.
5. Najmenší koreň $x^3 - 4.5x^2 + 1.3x + 8 = 0$. $[-1.08601]$
6. Koreň $\frac{1}{2}\sqrt{1-x^2} - x^2 = 0$.

   V cvičeniach 7 - 12 použite iteračnú schému na výpočet požadovaných koreňov

7. Určte $a^{\frac{1}{n}}$, kde $a > 0$ a n je celé číslo. Výsledok kontrolujte výpočtom $4^{\frac{1}{3}}$. $\left[x_{r+1} = \frac{1}{2}\left(x_r + \frac{a}{x_r^{n-1}}\right)\right]$
8. Nájdite korene $x^2 + 4x + 1 = 0$ a výsledok skontrolujte výpočtom koreňov kvadratickej rovnice.
9. Nájdite všetky tri korene $x^3 - 4.3x^2 + 1.4x + 7.8 = 0$. $[-1.08090, 2.54109, 2.83981]$
10. Nájdite kladné korene $\sin x - \frac{1}{2}x = 0$.
11. Nájdite kladné korene $x^2 - 2\sinh x + 1 = 0$. [0.67567]
12. Nájdite kladné korene $x^2 + 2\ln x - 4 = 0$.

    V cvičeniach 13 - 18 použite Newtonovu metódu a nájdite požadovaný koreň

13. Nájdite $23^{\frac{1}{3}}$ hľadaním núl rovnice $f(x) = 23 - x^3$. [2.84387]
14. Nájdite najmenší kladný koreň $\operatorname{tg} x + 2\tanh x = 0$.
15. Nájdite najväčší koreň $x^4 - 4x^3 + x^2 + 1.2 = 0$. [3.70665]
16. Nájdite najmenší koreň $x^4 - 3x^3 + 2x^2 - 3x - 1.6 = 0$.
17. Nájdite koreň $3x - e^{-x} = 0$.
18. Nájdite koreň $1 + \tanh x - 2\operatorname{tg} x = 0$.

Chapter 3 INTERPOLÁCIA A EXTRAPOLÁCIA.

Niekedy je spojitá funkcia $f(x)$ známa v tvare množiny diskrétnych hodnôt $y_i = f(x_i)$ v množine argumentov $x_1, x_2, ..., x_n$ kde $x_1 < x_2 < \cdots < x_n$. Ak sa toto stane, často je nutné odhadnúť hodnotu $f(\alpha)$ keď α leží medzi dvomi známymi argumentami x_i. Tento proces sa nazýva interpolácia funkcie $f(x)$ medzi jej známymi hodnotami a interpolovaná hodnota $f(\alpha)$ sa odhaduje použitím niekoľkých alebo všetkých hodnôt y_i. Rôzne metódy sú vhodné na interpoláciu, ale pokiaľ neurobíme nejaké predpoklady o funkcii, o chybe nemožno nič povedať. Ako všeobecné pravidlo chyba sa najlepšie redukuje výberom metódy reflektujúcej zdanliné zmeny $f(x)$. Niektoré faktory, ktoré môžeme brať do úvahy pri výbere interpolačnej metódy: či je $f(x)$ konvexná alebo konkávna pre $x_1 < x < x_n$, alebo či je oscilatorická, alebo či má ostrú krivosť v bode alebo bodoch intervalu. Odhad $f(\alpha)$, keď α leží mimo intervalu, alebo vľavo od x_1 alebo vpravo od x_n, sa nazýva extrapoláciou funkcie $f(x)$ a ak je proces náchylný k značným chybám musí sa používať opatrne. Ako v prípade interpolácie nič nemožno povedať o chybe extrapolácie, pokiaľ nevieme, alebo nepredpokladáme o funkcii nejaké všeobecné vlastnosti. Extrapolácia sa používa častejšie, ako by sme si mohli myslieť. Napríklad v Newtonovej metóde keď krivosť v bode nahradíme jej dotyčnicou, ktorá je potom rozšírená (extrapolovaná) pokiaľ nepretne os o_x, podobne numerické riešenie ODR o ktorom ešte budeme hovoriť.

Lineárna Interpolácia.

Nech dáta $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ patria neznámej hladkej funkcii $y = f(x)$ sú načrtnuté do grafu. Potom najjednoduchšia cesta odhadnúť hodnotu $y(x)$, keď x leží v intervale $x_i < x < x_{i+1}$ je spojiť body (x_i, y_i) a (x_{i+1}, y_{i+1}) rovnou úsečkou a potom použiť bod na úsečke s argumentom x ako aproximáciu $y(x)$. Toto je klasický proces a nazýva sa lineárna interpolácia ako na obrázku, kde A je bod (x_i, y_i), B je bod (x_{i+1}, y_{i+1}), a úsečka $\overline{AB}$ má rovnicu $y = \widetilde{y}(x)$. Potom v lineárnej interpolácii bod P na úsečke $\overline{AB}$ je použitý ako aproximácia bod Q na krivke $y = f(x)$. Jednoduchý výpočet ukayuje, že úsečka $y = \widetilde{y}(x)$ reprezentujúca lineárnu interpoláciu funkcie medzi dvomi bodmi (x_i, y_i) a (x_{i+1}, y_{i+1}) je daná:

$$\widetilde{y}(x) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i) + y_i, \quad \text{pre} \quad x_i < x < x_{i+1}. \tag{(4)}$$

Ak x vyberieme tak, aby buď $x < x_1$ alebo $x > x_n$, výsledok (4) bude lineárna extrapolačná formula pre $y = f(x)$ mimo intervalu $x_1 < x < x_n$. Výsledok (4) je vhodný pre interpoláciu keď variácia x_i and y_i medzi medzi susediacimi bodmi je malá, ale pretože formula vnáša chybu, vďaka jej zlyhaniu, ak by sme vzali do úvahy krivosť krivky, chyba by bola väčšia ak výsledok použijeme pre extrapoláciu.

Lagrangeova interpolácia.

Ak miesto lineárnej interpolácie, ktorá berie do úvahy postupne páry dát bodov $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, je možné, že lepší výsledok by sme dostali ak skonštruujeme polynóm $y = P(x)$, ktorý prechádzaa každým dátovým bodom. Pretože

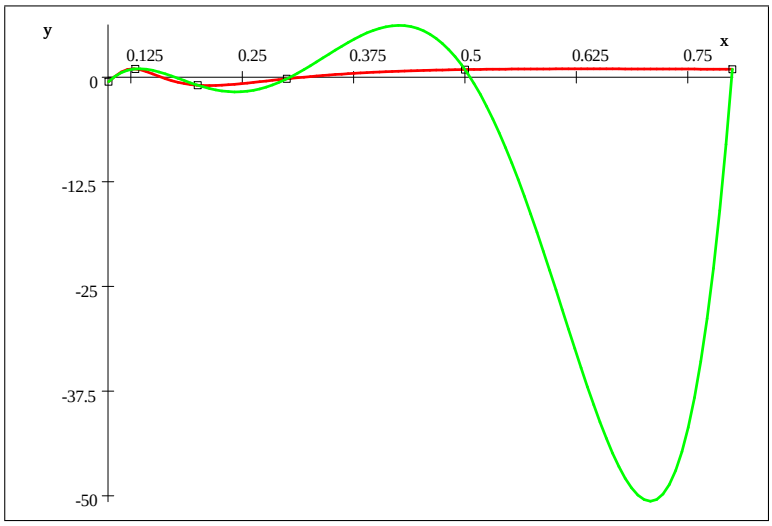
polynóm je hladkou krivkou možno predpokladať, že bude uvažovať aj krivosť funkcie, ku ktorej patria dáta funkcie a tak dostať lepšiu interpoláciu. V Lagrangeovej interpolácii polynóm $P(x)$ sa zvykne brať do úvahy polynóm s najmenším možným stupňom, ktorý prechádza všetkými dátovými bodmi, teda ak máme n dátových bodov polynóm bude najviac $(n-1)$ −vého stupňa. Polynóm je jediný, pretože n rovníc pre jeho n koeficientov možno nájsť tak, že budeme požadovať, aby prechádzal cez každý z n dátových bodov. Graf tohto polynómu cez interval $x_1 \leq x \leq x_n$ sa potom používa ako aproximácia neznámej funkcie $y = f(x)$ z ktorej sa predpokladá, že dátové body možno odvodiť z predpokladu, že $y = f(x)$ neukazuje veľké variácie ako sa x mení medzi postupnými argumentami $x_1, x_2, ..., x_n$ dátových bodov. Polynóm $y = P(x)$ je daný:

$$P(x) = \sum_{k=1}^{n} L_k(x)\, y_k, \qquad ((5))$$

kde

$$L_k(x) = \frac{(x-x_1)(x-x_2)\ldots(x-x_{k-1})(x-x_{k+1})\ldots(x-x_n)}{(x_k-x_1)(x_k-x_2)\ldots(x_k-x_{k-1})(x_k-x_{k+1})\ldots(x_k-x_n)}$$

má vlastnosť, ktorú sme požadovali, pretože je stupňa $(n-1)$ a prechádza každým dátovým bodom, teda definuje interpolačnú formulu pre interval $x_1 \leq x \leq x_n$. Polynómy $L_k(x)$, sa nazývajú fundamentálne Lagrangeove interpolačné polynómy, sú stupňa $(n-1)$, ale lineárna kombinácia tvoriaca $P(x)$ obsahujúca dáta body môže mať aj menší stupeň. To, že $L_k(x)$ má požadované vlastnosti ľahko vidieť z faktu, že pre $x = x_k$ každý $L_r(x_k)$ s $r = k$ obsahuje nulový faktor v čitateli, teda $L_r(x_k) = 0$, ale ak $r = k$ máme $L_k(x_k) = 1$, odkiaľ vidíme, že $P(x_k) = y_k$. Polynóm $P(x)$ zachováva požadované Lagrangeove interpolačné formuly pre množinu n data bodov $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Ak $n = 2$ výsledok (5) sa redukuje na lineárnu interpoláciu, ak $n = 3$ je to kvadratická, ktorá fituje parabolu cez tri body. Parabola je hladká krivka, so stabilnou zmenou gradientu, teda pretože berie do úvahy krivosť neznámej funkcie $y = f(x)$ cez tri požadované body, očakávame, že bude lepšou aproximáciou ako lineárna interpolácia. Avšak je nevhodné použiť Lagrangeovu interpoláciu pre viac ako tretí stupeň, pretože ak polynóm stupňa $(n-1) \gg 1$ je nútený prechádzať cez množinu n pevných bodov je to obvykle polynóm s veľkými osciláciami medzi susednými pármi dátových bodov, aj keď body samotné neindikujú takéto chovanie originálnej funkcie. Táto nevhodná charakteristika Lagrangeovho interpolačného polynómu vyššiehostupňa môže byť ilustrovaná konštrukciou interpolačného polynómu piateho stupňa pre funkciu $y(x) = \sin\left(\frac{1}{x}\right)$,



v intervale $0,1 \leq x \leq 0,8$. Ak skonštruujeme interpolačnú funkciu, presné extrémy funkcie, ktoré sú všeobecne známe sa zmenia na neurčité, ak vyberieme napríklad šesť bodov na grafe $y(x)$: $(0.1, -0.544021)$, $(0.13, 0.986959)$, $(0.2, -0.958924)$, $(0.3, -0.190568)$, $(0.5, 0.909297)$, $(0.8, 0.948985)$. Lagrangeov interpolačný polynóm prechádzajúci cez týchto šesť bodov je polynóm $p(x) = -47.953442 + 1039.947347x - 7963.493901x^2 + 26828.578780x^3 - 39901.683910x^4 + 21121.453960x^5$. Extrémne oscilácie je vidieť na obrázku. V tomto prípade namiesto šiestich dátových bodov by bolo lepšie použiť tri za sebou idúce Lagrangeove interpolácie na intervaloch $0.1 \leq x \leq 0.2$, $0.2 \leq x \leq 0.5$, a $0.3 \leq x \leq 0.8$, s posledným interpolačným polynómom použitým iba v intervale $0.5 \leq x \leq 0.8$. Avšak takáto zložená interpolačná schéma by pravdepodobne mala spojitú deriváciu v bodoch $x = 0.2$ and $x = 0.5$ kde sa stretajú parabolické aproximácie. Konštatujeme teda, že hlavný je teoretický prínos použitia v súvislosti s efektívnymi numerickými technikami.

Kubické splajny (spline).

Dôležité použitie interpolačných techník sa vyskytuje najmä v inžinierskom designe a všade, kde je nutné generovať hladké krivky s neznámymi funkciami, ktoré prechádzajú cez množinu dátových bodov bez oscilácií medzi nimi. Navrhnutý prístup je motivovaný starou inžinierskou drážkovacou technikou, ktorá produkuje také krivky trasovaním (obkresľovaním) podľa tenkého kovového pásika nazývaného spline, ktorý pôsobením tlaku v bodoch po jej dĺžke bol nútený prechádzať cez každý dátový bod. Je jasné, že Lagrangeova polynomická interpolácia je nevhodná, pretože produkuje oscilácie a v praxi máme veľa dátových bodov. Volíme taký prístup, že krivku aproximujeme po častiach polynómami 3 stupňa na každom z intervalov $x_i \leq x \leq x_{i+1}$ takým spôsobom, že prvá aj druhá derivácia krivky v koncových bodoch intervalu odpovedá aproximáciám zľava v x_i a aproximáciám sprava v x_{i+1}.Zložená aproximácia tohto typu sa nazýva aproximácia kubickými splineami. V matematickom prístupe k určeniu splinovej funkčnej aproximácie cez n dátových bodov (x_1, y_1) ,) (x_2, y_2) , . . . , (x_n, y_n) , x_i sa nazývajú nody (nuly!) aproximácie a odpovedajúce body y_i, kde sa priľahlé krivky stretávajú sa nazývajú uzly aproximácie. Matematické požiadavky, ktoré má spĺňať splineová aproximačná funkcia sú:

(a) Každá krivka prechádzajúca priľahlými bodmi (x_i, y_i) a (x_{i+1}, y_{i+1}) je ku-

bická.

(b) Zložená krivka po celom intervale musí interpolovať dáta prechádzajúc cez každý uzol.

(c) Krivka samotná a prvá a druhá derivácia zloženej krivky musí byť spojitá v uzloch x_i.

(d) V koncových bodoch x_1 a x_n intervalu musia byť predpísané podmienky v závislosti od toho, či dáta body indikujú, či medzi týmito bodmi má extrapolačná krivka pripomínať priamu čiaru, parabolu, alebo ukazovať iné chovanie: napríklad periodicitu na intervale $x_1 \leq x \leq x_n$. Kvôli podmienkam (a) - (c) sa druhá derivácia $f''(x)$ musí meniť lineárne na každom intervale $x_i \leq x \leq x_{i+1}$ a byť spojitá pozdĺž každého nodu, tak použitím Lagrangeovej interpolačnej formuly môžme písať

$$f''(x) = \frac{x_{i+1} - x}{x_{i+1} - x_i} f''(x_i) + \frac{x - x_i}{x_{i+1} - x_i} f''(x_{i+1}), \quad \text{pre} \quad x_i \leq x \leq x_{i+1}. \qquad ((6))$$

Integráciou tohto výledku dvakrát vzhľadom ku x máme

$$f(x) = \frac{1}{6} \frac{3x_{i+1}x^2 - x^3}{x_{i+1} - x_i} f''(x_i) + \frac{1}{6} \frac{x^3 - 3x_i x^2}{x_{i+1} - x_i} f''(x_{i+1}) + ax + b, \qquad ((7))$$

pre $x_i \leq x \leq x_{i+1}$, kde a a b sú ľubovoľné integračné konštanty. Aby $f(x)$ prechádzala bodmi (x_i, y_i) a (x_{i+1}, y_{i+1}), substitúciou týchto dvoch podmienok do vzťahu (7) určíme a, b, položíme $d_i = x_{i+1} - x_i$ dostaneme

$$\begin{array}{c} f(x) = \frac{1}{6d_i} \left[(x_{i+1} - x)^3 f''(x_i) + (x - x_i)^3 f''(x_i) \right] + \\ + \frac{1}{6d_i} \left[6y_i - d_i^2 f''(x_i) \right] (x_{i+1} - x) + \frac{1}{6d_i} \left[6y_{i+1} - d_i^2 f''(x_{i+1}) \right] (x - x_i), \end{array} \quad . \qquad ((8))$$

pre $x_i \leq x \leq x_{i+1}$. Na ďalšie pokračovanie musíme nájsť podmienky určujúce derivácie $f''(x_i)$ a $f''(x_{i+1})$, a toto môžeme dostať použitím doteraz nepoužitej podmienky, že prvá derivácia $f'(x)$ musí byť spojitá pozdĺž každého nodu. Aplikácia tejto podmienky znamená diferencovať (8) a požadovať aby derivácia keď $x = x_{i+1}$ v $i-$tom intervale, t.j., v jeho pravom koncovom bode sa rovnala derivácii keď $x = x_{i+1}$ v $(i+1)-$vom intervale, odpovedajúca jeho ľavému krajnému bodu odkiaľ dostaneme

$$d_{i-1} f''(x_{i-1}) + 2(d_{i-1} + d_i) f''(x_i) + d_i f''(x_{i+1}) = Y_i, \qquad ((9))$$

kde

$$Y_i = 6 \left(\frac{y_{i+1} - y_i}{d_i} - \frac{y_i - y_{i-1}}{d_{i-1}} \right). \qquad ((10))$$

Výsledok (9) je množina $n - 2$ lineárnych rovníc pre n derivácií $f''(x_i)$, a ak ich poznáme, potom spline aproximačnú funkciu tvorenú funkciami (8) vytvoríme (skonštruujeme) postupne na intervaloch $x_i \leq x \leq x_{i+1}$ s $i = 1, 2, ..., n-1$. Pre praktické použitie splinov je nutné, aby lineárny systém rovníc bol nesingulárny a aby sme mali extrémne účinný algoritmus ich riešenia. Pretože hodnoty $f''(x_1)$ a $f''(x_n)$ nemožno nájsť z podmienky, že $f''(x)$ je spojitá uzly x_1 a x_n pre tieto hodnoty musíme špecifikovať ako dodatočné podmienky. Výber hodnôt $f''(x_1)$ a $f''(x_n)$ sa robí intuitívne a je založený na spôsobe ako data body indikujú aby sa

interpolované krivky chovali (boli extrapolované) za koncovými bodmi intervalu $x_1 \leq x \leq x_n$. Typické sú tri výbery prirodzené alebo lineárne spline koncové podmienky, parabolické spline koncové podmienky, periodické spline end podmienky.

Prirodzené, alebo lineárne koncové podmienky.
V tomto prípade volíme koncové podmienky

$$f''(x_1) = f''(x_n) = 0. \qquad ((11))$$

Tieto podmienky sa nazývajú lineárne koncové podmienky, pretože aj keď sa vnútri intervalu používajú kubické polynómy, nulová druhá derivácia v $x = x_1$ a $x = x_n$ spôsobí, že aproximácia bude lineárne na koncoch intervalu.

Parabolické splajny koncové podmienky.
Tento výber koncových podmienok obsahuje

$$f''(x_1) = f''(x_2) \quad \text{a} \quad f''(x_{n-1}) = f''(x_n). \qquad ((12))$$

Tieto podmienky sa nazývajú parabolické spline koncové podmienky, pretože ich dôsledkom je, že $f''(x)$ je konštantná v každom koncovom intervale, čo spôsobí, že kubická interpolačná formula sa redukuje na kvadratickú alebo parabolickú aproximáciu.

Periodické splajny koncové podmienky.
Ak sú dáta periodické na intervale $x_1 \leq x \leq x_n$, potom sú vhodné nasledujúce koncové podmienky

$$f(x_1) = f(x_{n-1}) \quad \text{a } f'(x_n) = f'(x_2). \qquad ((13))$$

Aj iné koncové podmienky sú možné napríklad lineárne spline koncové podmienky sú aplikované na jednom konci a parabolické koncové spline podmienky na druhom konci. Jedna koncová podmienka, ktorá je dôležitejšia než parabolická koncová podmienka je podmienka vedúca ku úplnému kubickému splinu, napríklad spline, ktorý interpoluje $f'(x)$ ako aj $f(x)$ v oboch koncoch x_1 a x_n. Tento spline má vyšší rád konvergencie ak maximum dĺžky kroku sa blíži ku nule, a je často implementovaný použitím lokálnych aproximácií derivácií, ktoré zachovávajú vyšší rád konvergencie.

Funkcia $y(x) = \sin\left(\frac{1}{x}\right)$ je na obrázku ako červená krivka, kde je aj príklad splineovej aproximácie is superimposed kubické spline aproximácia s prirodzenými okrajovými podmienkami. Šesť interpolačných data bodov sú ukázané ako body.

Cvičenia.

Cvičenia v tomto odseku vyžadujú použiť computer.

1. Načrtnite graf funkcie $f(x) = \frac{x}{(1+x^2)}$ v intervale $0 \leq x \leq 3$. Vyberte štyri body na grafe a po konštrukcii polynómu, ktorý prechádza cez všetky tieto body výsledok porovnajte s originálnou funkciou.

2. Načrtnite graf funkcie $f(x) = \frac{\sin x}{(1+x^2)}$ v intervale $0 \leq x \leq \pi$. Vyberte štyri body na grafe a po konštrukcii polynómu, ktorý prechádza cez všetky tieto body výsledok porovnajte s originálnou funkciou.

3. Načrtnite graf funkcie $f(x) = 1 + x \sin x$ v intervale $0 \leq x \leq 2\pi$. Vyberte sedem bodov na grafe a po konštrukcii polynómu, ktorý prechádza cez všetky tieto body výsledok porovnajte s originálnou funkciou. Potvrďte aproximáciu výberom iných sedem bodov.

4. Načrtnite graf funkcie $f(x) = (1 - x^5)^{\frac{1}{5}}$ v intervale $0 \leq x \leq 1$. Vyberte sedem bodov na grafe a po konštrukcii polynómu, ktorý prechádza cez všetky tieto body výsledok porovnajte s originálnou funkciou. Potvrďte aproximáciu výberom iných sedem bodov.

5. Načrtnite graf funkcie $f(x) = 1 - 2x \cos x$ v intervale $0 \leq x \leq 2\pi$. Vyberte sedem bodov na grafe a skonštruujte spline aproximatívnu funkciu v intervale $0 \leq x \leq 2\pi$ použitím parabolických splineov koncových podmienok. Načrtnite graf spline funkcie a porovnajte ho s grafom originálnej funkcie. Opakujte výpočet použitím linear spline funkcie koncové podmienky a porovnajte s predošlým grafom.

6. Načrtnite graf funkcie $f(x) = (1 - x^7)^{\frac{1}{7}}$ v intervale $0 \leq x \leq 1$. Vyberte sedem bodov na grafe a skonštruujte spline aproximatívnu funkciu v intervale $0 \leq x \leq 1$ použitím lineárnych splineov koncových podmienok. Načrtnite graf spline funkcie a porovnajte ho s grafom originálnej funkcie. Opakujte výpočet použitím parabolických spline funkcie koncové podmienky a porovnajte s predošlým grafom.

Chapter 4 NUMERICKÁ INTEGRÁCIA.

Numerická integrácia, nazývaná tiež numerická kvadratúra, sa používa keď buď potrebujeme vyčísliť určitý integrál a nevieme ho vypočítať analyticky, alebo ak sa vo vyčíslení integrálu vyskytujú špeciálne funkcie a analytické riešenie je príliš komplikované pre praktické použitie. Typickým príkladom, ktorý možno vypočítať numericky je

$$I = \int_0^5 \frac{\sin 3x}{\sqrt{x^2 + x + 1}} dx,$$

ktorého hodnota je $I = 0.364873$. Ukážeme tri rôzne numerické integračné schémy na výpočet určitých integrálov a konkrétne lichobežníkové pravidlo, Simpsonovo pravidlo a Gaussovu integráciu. Z týchto metód je prvá najmenej presná, zatiaľ čo posledná vyžaduje vysokú presnosť s niekoľkými výpočtovými krokmi na rozdiel od často používaného Simpsonovho pravidla.

Lichobežníkové pravidlo.

Jeho základom je veľmi jednoduché pravidlo, ktoré sa dá ľahko pochopiť z obrázku,

kde integrál $I = \int_a^b f(x)dx$ aproximujeme plochou lichbežníka $PQRS$ na intervale $a \leq x \leq b$ súvisiacim s grafom $y = f(x)$. Lichobežníková aproximácia $I = \int_a^b f(x)dx$. Plošný obsah obrazca $PQRS = \frac{1}{2}(b-a)[f(a) + f(b)]$, a aproximácia určitého integrálu je daná is given by b a

$$\int_a^b f(x)dx \approx \frac{1}{2}(b-a)[f(a) + f(b)]. \qquad ((14))$$

Ak položíme $b - a = h$, a chybu aproximácie určitého integrálu jednoduchým lichobežníkovým pravidlom označíme $E(h)$, máme

$$E(h) = \frac{1}{2}(b-a)[f(a) + f(b)] - \int_a^b f(x)dx,$$

potom

$$\int_a^b f(x)dx = \frac{1}{2}(b-a)[f(a) + f(b)] - E\,(h) \qquad ((15))$$

aproximáciu (14) nahradíme presným výsledkom (15). Existujú rôzne spôsoby odvodenia (14) napríklad použitie lineárnej interpolácie na reprezenáciu $y(x)$ medzi $x = a$ a $x = b$, a potom integrovať výsledok. Aj keď presná hodnota chyby $E(h)$ nie je známa výraz pre chybu možno odvoditž predpokladu, že $f(x)$ je vhodne

diferencovateľná v integračnom intervale $a \leq x \leq b$. Chybu iba skonštatujeme, pretože jej odvodenie je podobné ako pre presnejšie Simpsonovo pravidlo, ktoré bude odvodené neskôr. Tak máme $E(h) = \frac{1}{12}h^3 f''(\xi)$, pre nejaké ξ, $a \leq \xi \leq b$. Potom (15) bude

$$\int_a^b f(x)dx = \frac{1}{2}(b-a)[f(a)+f(b)] - \frac{1}{12}h^3 f''(\xi), \qquad ((16))$$

kde $a \leq \xi \leq b$. Lepší odhad určitého integrálu $\int_a^b f(x)dx$ možno získať, delením intervalu $a \leq x \leq b$ na n podintervalov a aplikáciou odhadu (16) na každom z nich a potom sčítať výsledok. Často sa používa delenie na n rovnakých intervalov s dĺžkou $h = \frac{(b-a)}{n}$, kde h nazývame dĺžkou kroku. Ak teda položíme $x_i = a + ih$, pre $i = 0, 1, ..., n$, dostaneme tzv zložené lichobežníkové pravidlo

$$\int_a^b f(x)dx = \frac{1}{2}h\left[f(a) + 2\sum_{i=1}^{n-1} f(x_i) + f(b)\right] - \frac{1}{12}(b-a)h^2 f''(\eta), \qquad ((17))$$

kde $a \leq \eta \leq b$. Chybu zloženej lichobežníkovej metódy dostaneme z (16) dodaním chýb na jednotlivých intervaloch. Detaily ponecháme ako cvičenie. Aj keď η nie je známa, vždy je možné odhadnúť najväčšiu a najmenšiu hodnotu $f''(x)$ v $a \leq x \leq b$, a táto môže byť v odhade zloženej lichobežníkovej metódy ako $f''(\eta)$. V praktických applikáciách sa odhad chyby lichobežníkového pravidla používa obyčajne vtedy ak chceme ukázať, že ak počet subintervalov rastie, chyba aproximácie klesá na $\frac{(b-a)h^2}{12}$, kde $h = \frac{(b-a)}{n}$. Chybu často aproximujeme vytvorením dvoch aproximácií s rôznymi odhadmi použitím asymptotického chovania na odhad chyby výsledku odpovedajúcemu malému h. Iný prístup je porovnať výsledok s výsledkom Simpsonovej metódy.

Example 14 *Použitím lichobežníkového pravidla s $n = 10, 30$ a 50 podintervalov vypočítajte $I = \int_0^5 \frac{\sin 3x}{\sqrt{x^2+x+1}}dx$, a aproximujte chybu ak použijeme 50 podintervalov.*

Solution 15 *Výpočtom na počítači sme dostali:*

n	10	30	50
$I_{lichob(n)}$	0.290422	0.356897	0.362010

.

Výsledok pre $I_{lichob(50)}$ mal byť porovnaný s výsledkom, ktorý bol získaný metódou vyššieho rádu s výsledkom $I = 0.364873$ presným na 6 desatinných miest. Namiesto použitia $f''(\eta)$ pri aproximácii chyby s $n = 50$, kde η je známe využijeme ľahko vypočítateľnú strednú hodnotu f''_{av} funkcie $f''(x)$ na intervale, kde

$$f''_{av} = \frac{1}{b-a}\int_a^b f''(x)dx = \frac{1}{b-a}[f'(b) - f'(a)].$$

Máme $b-a=5$ *a veľkosť kroku* $h=\frac{5}{50}=0.1$, *t.j.*

$$f''_{av}=\frac{1}{5}\int_0^5 f''(x)dx=\frac{1}{5}[f'(5)-f'(0)]=-0.686.$$

Použitím f''_{av} *v odhade chyby miesto* $f''(\eta)$ *vedie ku* $\frac{1}{12}\cdot 5\cdot(0.1)^2\cdot(-0.686)=-0.002858$ *ako odhadu chyby. Teda ak uvažujeme túto chybu, potom odhad integrálu je* $0.362010-(-0.002858)=0.364868$. *Ak to porovnáme s výsledkom* $I=0.364873$ *vidíme, že chyba aproximácie odhadu je dobrá.* □

Simpsonovo pravidlo.

V najjednoduchšej forme je lichobežníkové pravidlo aplikované na výpočet $\int_a^b f(x)dx$ reprezentuje funkciu $f(x)$ jednoduchou lichobežníkovou plochou $PQRS$ ako na obrázku, kde v intervale $a\le x\le b$ funkciu $y=f(x)$ aproximuje úsečka QR. Presnejší výsledok je možné očakávať ak bod na krivke $y=f(x)$ je vybratý vnútri intervalu $a\le x\le b$ a $f(x)$ aproximujeme ako parabolu prechádzajúcu cez dva koncové body intervalu a vybraný vnútorný bod ako na obrázku. Položme $b=a+2h$, kde h je veľkosť kroku a dodatočný bod v integračnom intervale bude $x=a+h$, stred intervalu. Parabola, ktorá bude fitovať tieto tri body musí teda prechádzať postupne bodmi $(a,f(a))$, $(a+h,f(a+h))$, a $(a+2h,f(a+2h))$. Lagrangeova interpolačná kvadratická formula, ktorá fituje tieto tri body je:

$L(x)=\frac{1}{2}\frac{(x-a-h)(x-a-2h)}{h^2}f(a)-\frac{(x-a)(x-a-2h)}{h^2}f(a+h)+\frac{1}{2}\frac{(x-a)(x-a-h)}{h^2}f(a+2h)$.

Integráciou $L(x)$ cez interval $a\le x\le a+2h$ a zjednodušením výsledku máme

$$\int_a^{a+2h} f(x)dx\approx\frac{1}{3}h[f(a)+4f(a+h)+f(a+2h)],\qquad((19))$$

čo nazývame Simpsonovo pravidlo, alebo Simpsonovo tretinové pravidlo. Výsledok (19) možno zapísať pomocou koncových bodov integrálneho intervalu a a $b=a+2h$ ako

$$\int_a^b f(x)dx\approx\frac{1}{6}(b-a)\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right].\qquad((20))$$

Ak chybu v Simpsonovom pravidle označíme $E(h)$, aproximácia v (19) zameníme za presný výsledok

$$\int_a^{a+2h} f(x)dx=\frac{1}{3}h[f(a)+4f(a+h)+f(a+2h)]-E(h).\qquad((21))$$

Odvodíme výraz pre $E(h)$, predtým položíme $a=c-h$ a $b=c+h$. Pomocou c a h (21) prepíšeme

$$E(h)=\frac{1}{3}h[f(c-h)+4f(c)+f(c+h)]-\int_{c-h}^{c+h} f(x)dx.$$

Diferencujme podľa h a máme

$$E'(h) = \frac{1}{3}[f(c-h)+4f(c)+f(c+h)]+\frac{1}{3}h[-f'(c-h)+f'(c+h)]-[f(c+h)+f(c-h)],$$

Ak položíme $h = 0$, dostaneme $E'(0) = 0$. Differentiation of $E'(h)$ dostaneme

$$E''(h) = \frac{1}{3}[f'(c-h) - f'(c+h)] + \frac{1}{3}h[f''(c+h) + f''(c-h)],$$

položme $h = 0$ a máme $E''(0) = 0$. Ešte konečná derivácia $E'''(h) = \frac{1}{3}h[f'''(c+h) - f'''(c-h)]$, ale to možno zjednodušiť použitím Taylorovho rozvoja $f(c+h)$ so zvyškom po prvom člene $f'''(c+h) = f'''(c-h) + 2hf^{(4)}(\xi)$, kde ξ je neznáme, ale leží v intervale $c-h < \xi < c+h$. The error term can now be found by integrating this last result three times using the results $E''(0) = E'''(0) = 0$. Potom máme

$$\int_0^h E'''(t)dt = E''(h) - E''(0) = E''(h),$$

teda

$$E''(h) = \frac{2}{3}f^{(4)}(\xi)\int_0^h t^2dt = \frac{2}{9}h^3f^{(4)}(\xi).$$

Po ďalšej integrácii

$$\int_0^h E''(t)dt = E'(h) - E'(0) = E'(h)$$

dostaneme

$$E'(h) = \frac{2}{9}f^{(4)}(\xi)\int_0^h t^3dt = \frac{1}{18}h^4f^{(4)}(\xi).$$

Nakoniec po ďalšej integrácii dostaneme

$$E(h) = \frac{1}{18}f^{(4)}(\xi)\int_0^h t^4dt = \frac{1}{90}h^5f^{(4)}(\xi), \qquad ((22))$$

čo je požadovaný výraz pre chybu. Použitím tohto výsledku v (21) dáva

$$\int_a^{a+2h} f(x)dx = \frac{1}{3}h[f(a) + 4f(a+h) + f(a+2h)] - \frac{1}{90}h^5f^{(4)}(\xi). \qquad ((23))$$

Pretože $f^{(4)}(\xi)$ vystupuje ako faktor v $E(h)$, toto ukazuje dosť prekvapujúci fakt, že aj keď Simpsonovo pravidlo sme odvodili tak, že kvadratický polynóm prechádza tromi bodmi, pravidlo je presné pre kubické polynómy. Ak pre lichobežníkové pravidlo, presnosť Simpsonovho pravidla možno zlepšiť rastúcim počtom podintervalov, ale pretože pravidlo je ekvivalentné s konštrukciou paraboly cez tri ekvidištančné body, použitie pravidla cez viac než tri body počet bodov vybraných pre interval $a \le x \le b$ musí byť nepárne, teda počet intervalov musí byť párny. Delenie

intervalu $a \leq x \leq b$ na $2n$ rovnakých podintervalov každý s dĺžkou $h = \frac{(b-a)}{2n}$, a sčítaním výsledkov dostaneme zložené Simpsonovo pravidlo s chybou

$$\int_a^b f(x)dx = \frac{1}{3}h\left[f(a) + 4\sum_{i=1}^{n} f(a+(2i-1)h) + 2\sum_{i=1}^{n-1} f(a+2ih) + f(b)\right] - \\ -\frac{1}{180}(b-a)h^4 f^{(4)}(\eta), \tag{(24)}$$

kde η je neznáma, ale taká, že $a < \eta < b$. Chybový výraz v zloženom pravidle (24) dostaneme nasledovne. Nech $x_i = a + 2ih$, s $i = 0, 1, ..., n$, a nech ξ_i bude hodnota ξ v intervale $x_i \leq x \leq x_{i+1}$ vhodné pre Simpsonovo pravidlo aplikované na tento interval. Tak ak tvoríme zložené Simpsonovo pravidlo v každom z týchto intervalov sčítame. Teraz každá derivácia $f^{(4)}(\xi_i)$ musí vyhovovať podmienke

$$\min_{a\leq x\leq b} f^{(4)}(x) \leq f^{(4)}(\xi_i) \leq \max_{a\leq x\leq b} f^{(4)}(x),$$

potom sčítanie týchto n výsledkov nasledovaných delením n dáva

$$\min_{a\leq x\leq b} f^{(4)}(x) \leq \frac{1}{n}\sum_{i=1}^{n} f^{(4)}(\xi_i) \leq \max_{a\leq x\leq b} f^{(4)}(x).$$

Nakoniec predpokladáme, že $f^{(4)}(x)$ je spojitá, z vety o medzihodnote máme, že existuje nejaké η, $a < \eta < b$, také, že $f^{(4)}(\eta) = \frac{1}{n}\sum_{i=1}^{n} f^{(4)}(\xi_i)$. Ak použijeme výsledok $h = \frac{(b-a)}{2n}$, chybový výraz pre zložené Simpsonovo pravidlo bude

$$-\frac{1}{90}h^5\sum_{i=1}^{n} f^{(4)}(\xi_i) = -\frac{1}{180}(b-a)h^4 f^{(4)}(\eta).$$

Example 16 *19.6 Použitím zloženého Simpsonovho pravidla s $n = 10, 30$ a 50 podintervalov na výpočet*

$$I = \int_0^5 \frac{\sin 3x}{\sqrt{x^2+x+1}}dx$$

a porovnajte výsledok získaný s výsledkom $I = 0.364873$, ktorý je presný na šesť desatinných miest. Porovnajte výsledok integrácie tohto integrálu s lichobežníkovým pravidlom.

Solution 17 *Nasledujúce výsledky sme dosiahli pomocou počítača:*

n	10	30	50
$I_{simp(n)}$	0.376738	0.365019	0.364892

Porovnanie s výsledkom $I = 0.364873$, známom ako správny na šesť desatinných miest s $Isimp(50) = 0.364892$, ukazuje, že $I_{simp(50)}$ presahuje správny výsledok o 0.000025.

Keď porovnáme zložené Simpsonovo pravidlo so zloženým lichobežníkovým pravidlom mali by sme si spomenúť, že Simpsonovo pravidlo rozdelí interval integrovania do $2n$ podintervalov, zatiaľ čo zložené lichobežníkové pravidlo iba do n podintervalov. Nasledujúci computrový výsledok provide porovnanie na tomto základe.

n	20	40	60	80	100
$I_{lichob(n)}$	0.346825	0.360395	0.362886	0.363756	0.364158
$I_{simp(\frac{n}{2})}$	0.376738	0.365626	0.365019	0.364919	0.364892

Gaussova kvadratúra.

Okrem už zmienených metód, dôležitá je metóda C. F. Gaussa. Ten ukázal, že ak numericky počítame integrál v štandardnej forme $\int_{-1}^{1} f(x)dx$, body x_i v ktorých máme dané hodnoty integrandu $f(x)$ sú vybrané špeciálnym spôsobom, potom ak , použijeme vzorové (sample) body, výsledok bude presný v prípade, že $f(x)$ je ľubovoľný polynóm stupňa $2n-1$ alebo menší. Na rozdiel od Simpsonovho pravidla n vzorových bodov x_i je nerovnomerne rozdelených v integračnom intervale $-1 \leq x \leq 1$, a sú obsiahnuté vnútri intervalu.

Vzorové (testové) body, alebo nody ako im hovoríme sú vybrané tak, aby integračná formula pomocou ktorej presne zintegrujeme polynóm takého stupňa ako je možné. Ukazuje sa, že n testovacích bodov je reálnych a ležia v otvorenom intervale $(-1, 1)$ a polynómy stupňa $2n-1$ sa dajú zintegrovať presne.

Trochu odlišný prístup k integrovaniu zvolíme, ak špecifikujeme nejaké vzorové body, a potom sa pokúsime nájsť ostatné aby sme mohli integrovať polynómy čo najväčšieho možného stupňa. Formuly tohto typu, ktoré počítajú funkčné hodnoty v dvoch koncoch integračného intervalu sa volajú Lobattove formuly, a lichobežníkové a Simpsonovo pravidlo sú formuly najnižšieho rádu, ktoré patria do tejto triedy.

Myšlienka je, že ak je vhodné špecifikovať testové body v koncových bodoch integračného intervalu je možné postupovať týmto spôsobom. Avšak ako sa dá očakávať, ak sa tento prístup použije, nie je možné dostať takú presnú formulu ako v prípade ak nedávame obmedzenia na testvé body.

Predošlé argumenty sú založené na predpoklade, že funkcie sa aproximujú (algebraickými) polynómami, aj keď občas je prirodzenejšie aproximovať ich trigonometrickými polynómami (konečnými Fourierovými radmi).

Zložené lichobežníkové pravidlo je v skutočnosti optimálnou formulou typu Gaussovskej integrácie založenej na trigonometrickej aproximácii. Výsledkom je rýchlejšia konvergencia, než ľubovoľná mocnina h ak sa použijú na periodickú analytickú funkciu cez násobok poriódy, tak pre tento dôvod sa použije výpočet Fourierových koeficientov.

Aby sme ilustrovali prístup používaný na získanie integračnej formuly tohto typu uvažujme najjednoduchšiu situáciu, pre ktorú $n = 2$, t.j. použijeme iba dva testové body x_1 a x_2 s $-1 < x_1 < x_2 < 1$ a integračná formula bude mať tvar

$$\int_{-1}^{1} f(x)dx \approx w_1 f(x_1) + w_2 f(x_2).$$

V tejto etape sú hodnoty dvoch testovacích bodov x_1 a x_2 neznáme, podobne ako čísla w_1 a w_2, ktoré sa nazývajú váhy pre integračnú formulu v týchto testovacích bodoch. Na určenie týchto štyroch čísel zavedieme požiadavku, že táto formula bude presná, ak $f(x)$ je ľubovoľný polynóm stupňa $2n-1 = 3$, alebo menší. Nech $f(x)$ je kubický polynóm $f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$, v ktorom sú koeficienty c_0, c_1, c_2, c_3 ľubovoľné. Aby bola integracia presná, čísla x_1, x_2, w_1 a w_2 musia byť také, že

$$\begin{aligned}&\textstyle\int_{-1}^{1}(c_0 + c_1 x + c_2 x^2 + c_3 x^3)dx = \\ &= w_1 \left(c_0 + c_1 x_1 + c_2 x_1^2 + c_3 x_1^3\right) + w_2 \left(c_0 + c_1 x_2 + c_2 x_2^2 + c_3 x_2^3\right).\end{aligned}$$

Vypočítaním integrálu naľavo a porovnaním dostaneme:

$(koeficient\ \ c_0) : w_1 + w_2 = \int_{-1}^{1} dx = 2$

$(koefficient\ c_1) : w_1 x_1 + w_2 x_2 = \int_{-1}^{1} x dx = 0$

$(koefficient\ c_2) : w_1 x_1^2 + w_2 x_2^2 = \int_{-1}^{1} x^2 dx = \frac{2}{3}$

$(koefficient\ c_3) : w_1 x_1^3 + w_2 x_2^3 = \int_{-1}^{1} x^3 dx = 0.$

Existuje jediné rišenie: $x_1 = \frac{-1}{\sqrt{3}}, x_2 = \frac{1}{\sqrt{3}}, w_1 = 1$ a $w_2 = 1$. Teda ak $n = 2$, máme:

Testovacie body $x_1 = -\frac{1}{\sqrt{3}}, x_2 = \frac{1}{\sqrt{3}}$,

váhy $w_1 = 1, w_2 = 1$, teda extrémne jednoduchá dvojbodová integračná formula dáva presný výsledok ak $f(x)$ je polynóm stupňa 3 alebo menej, potom $\int_{-1}^{1} f(x)dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$.

Ak tento prístup rozšírime na n bodov, skúmanie derivácie formuly ukazuje že testové body $x_1, x_2, ..., x_n$ sú jednoducho n koreňov Legendrovho polynómu $P_n(x) = 0$ stupňa n, s odpovedajúcimi váhami w_i v x_i danými $w_i = 2\frac{[P'(x_i)]^2}{(1-x_i^2)}$, pre $i = 1, 2, ..., n$. Všeobecná integračná formula zahrňujúca n bodov bude

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

a tieto výsledky sa nazývajú Gaussova integračná formula, alebo Gauss–Legendrove integračné formuly. Možno ukázať, že zvyšok Gauss–Legendrových integračných formúl, ktorý musíme dodať na pravú stranu posledného výsledku, aby bol presný pre každú funkciu $f(x)$ so spojitou deriváciou $f^{(2n)}(x)$ je $R_n = \frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3} f^{(2n)}(\xi)$, pre nejaké neznáme ξ, také že $-1 < \xi < 1$. Zoznam Gaussových testovacích bodov x_i a ich odpovedajúce váhy w_i je daná v tabuľke, pre $n = 2, 3, 4, 5, 10, 16$. Ako by sme mohli očakávať ak $f(x)$ je ľubovoľný polynóm stupňa $2n - 1$ chyba výrazu v Gaussovej integrácii je $R_n \equiv 0$, potvrdzujúca, že v tomto prípade je výsledok presný.

Gaussovské Sampling body a váhy.

n	i	x_i	w_i
2	1	−0.5773502692	1.0000000000
	2	0.5773502692	1.0000000000
3	1	−0.7745966692	0.5555555556
	2	0.0000000000	0.8888888889
	3	0.7745966692	0.5555555556
4	1	−0.8611363115	0.3478548451
	2	−0.3399810436	0.6521451548
	3	0.3399810436	0.6521451548
	4	0.8611363115	0.3478548451
5	1	−0.9061798459	0.2369268851
	2	−0.5384693101	0.4786286705
	3	0.0000000000	0.5688888889
	4	0.5384693101	0.4786286705
	5	0.9061798459	0.2369268851
10	1	−0.9739065285	0.0666713443
	2	−0.8650633667	0.1494513492
	3	−0.6794095683	0.2190863625
	4	−0.4333953941	0.2692667193
	5	−0.1488743390	0.2955242247
	6	0.1488743390	0.2955242247
	7	0.4333953941	0.2692667193
	8	0.6794095683	0.2190863625
	9	0.8650633667	0.1494513492
	10	0.9739065285	0.0666713443
16	1	−0.9894009350	0.0271524594
	2	−0.9445750231	0.0622535230
	3	−0.8656312024	0.0951585117
	4	−0.7554044084	0.1246289713
	5	−0.6178762444	0.1495959888
	6	−0.4580167777	0.1691565194
	7	−0.2816035508	0.1826034150
	8	−0.0950125098	0.1894506105
	9	0.0950125098	0.1894506105
	10	0.2816035508	0.1826034150
	11	0.4580167777	0.1691565194
	12	0.6178762444	0.1495959888
	13	0.7554044084	0.1246289713
	14	0.8656312024	0.0951585117
	15	0.9445750231	0.0622535239
	16	0.9894009350	0.0271524594

Zdanlivé obmedzenie integrálu na štandardný interval $-1 \leq x \leq 1$ nie je dôležité, pretože ak interval v integrále je $I = \int_a^b f(x)dx$, kde a a b sú konečné, jednoduchá zámena premenných $x = \frac{1}{2}(b+a) + \frac{1}{2}(b-a)u$ zmení integrál na $I =$

$\frac{b-a}{2}\int_{-1}^{1} F(u)du$, kde $F(u)$ je funkcia $f(x)$ po zámene premennej. Presnosť získame, ak použijeme $n-$bodovú Gaussovu integračnú formulu závisiacu od rozsahu, v ktorom možno integrand aproximovať polynómom stupňa $2n-1$. Aby sme ilustrovali podstatu, aplikujeme päťbodovú formulu na nasledujúci integrál, pre ktorý vieme analytické riešenie a použijeme ho na porovnavie:

$$I = \int_0^{\frac{1}{2}} \frac{dx}{(1-x^2)^{\frac{1}{2}}} = \arcsin\left(\frac{1}{2}\right) = \frac{\pi}{6} = 0.523599.$$

Zámena premennej $x = \frac{1}{4}(1+u)$ zobrazuje interval $0 \leq x \leq \frac{1}{2}$ na interval $-1 \leq u \leq 1$, po tejto substitúcii

$$I = \int_{-1}^{1} \frac{du}{(15-2u-u^2)^{\frac{1}{2}}}.$$

Položíme $f(u) = \frac{1}{(15-2u-u^2)^{\frac{1}{2}}}$ a aplikujeme päťbodovú Gaussovu formulu a dostaneme

$$I \approx 0.236927f(-0.906180) + 0.478629f(-0.538469) + 0.568889f(0)+$$

$$+0.478629f(0.538469) + 0.236927f(0.906180) = 0.523599.$$

V tomto prípade numerickej aproximácie vidíme presnosť na šesť desatinných miest.

Kľúčovou myšlienkou v modernej integrácii je adaptívny algoritmus. Znamená to, že chyba integrálu vypočítaná na intervale sa aproximuje porovnaním s výsledkom dosiahnutým iným prístupom. Tak chyba lichobežníkového pravidla môže byť odhadnutá porovnaním s výsledkom dosiahnutým Simpsonovým pravidlom. Ak výsledok nie je dostatočne presný interval rozdelíme na polovicu a dva intervaly sú skúmané oddelene. Redukcia dĺžky intervalu významne redukuje chybu. Efekt rozpolenia intervalu redukuje chybu v polovici intervalu približne na osminu pôvodnej chyby, teda pretože operácia integrovania je lineárna , chyba na celom intervale sa redukuje asi štyrikrát. Ak tento argument rozšírime, a interval rozdelíme na mnoho častí, presné hodnoty sa na jednotlivých častiach sčítajú a na celom intervale dostaneme presnejší výsledok s chybou takou istou ako na jednotlivom parciálnom intervale. Pri tomto prístupe sa aplikujú dve formuly s čo najväčším počtom bodov. Tento prístup je výpočtovo účinný ak sa použije kombinácia lichobežníkového a Simpsonovho pravidla čo vidieť z faktu, že na výpočet chyby je nutný iba jeden extra výpočet. Moderné prístupy využívajú Gaussovo pravidlo vysokého stupňa ako základnú formulu a špeciálnu formulu omnoho vyššieho rádu , ktorá používa toľko výpočtov funkcií koľko je možných pre odhad chyby.

Cvičenia.

Cvičenia v tomto odseku vyžadujú použiť computer.

1. Použite zložené Simpsonovo pravidlo s krokom dĺžky $h = 0.5$ na výpočet $I = \int_1^3 (2x^3 - 3x^2 + 4x - 1)dx$, a potom verifikujte presným výpočtom.

2. Použite zložené lichobežníkové pravidlo s krokom dĺžky $h = 0.1$ na výpočet $I = \int_0^1 \frac{dx}{1+x^2}$, a odhadnite chybu. Porovnajte s presným výsledkom $I = \frac{\pi}{4}$. Zopakujte výpočet použitím zloženého Simpsonovho pravidla s rovnakým krokom ale bez odhadu chyby.

3. Použite zložené lichobežníkové a Simpsonovo pravidlo, každé s 10 podintervalmi na odhad $I = \int_0^\pi \frac{\sin x}{x} dx$, a porovnajte váš výsledok s výsledkom $I = 1.851937$, presným na šesť desatinných miest.

4. Použite zložené lichobežníkové a Simpsonovo pravidlo, každé s krokom dĺžky $h = 0.2$, na odhad $I = \int_0^2 x^2 e^{-x} dx$, a porovnajte váš výsledok s analytickým riešením $I = \frac{1}{13} + \frac{1}{2}\operatorname{arctg} 5 - \frac{1}{8}\pi$.

5. Použite zložené Simpsonovo pravidlo s dĺžkou kroku $h = 0.4$ na odhad $I = \int_2^6 \frac{\ln(2+3\sqrt{x})}{1+x^2} dx$, a porovnajte svoj výsledok s výsledkom $I = 0.596545$, ktorý je presný na šesť desatinných miest.

6. Použite zložené lichobežníkové a Simpsonovo pravidlo, každé s krokom dĺžky $h = 0.2$, na odhad $I = \int_0^4 \left(1 - \frac{x}{4}\right)^4 x^{\frac{1}{2}} dx$. Porovnajte váš výsledok s presným riešením, ktoré plynie zo všeobecného výsledku

$$I(z,n) = \int_0^n \left(1 - \frac{x}{n}\right)^n x^{z-1} dx = \frac{1 \cdot 2 \cdot 3 \ldots n}{z(z+1)(z+2)\cdots(z+n)} n^z,$$

ktorý plynie z definície gamma funkcie $\lim_{n\to\infty} I(z,n) = \Gamma(z)$. Vysvetlite, prečo zámenou 4 za 50 v originálnom integrále a výpočte výsledku použitím zloženého Simpsonovho pravidla s viacerými deleniami pravdepodobne nevedie k zlepšeniu úbohého odhadu, ktorý it provides $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$.

7. Besselova funkcia $J_1(x)$ má integrálnu reprezentáciu $J_1(x) = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sin(x \cos\Theta) \cos\Theta d\Theta$. Použitím zloženého Simpsonovho pravidla s krokom dĺžky $h = \frac{\pi}{20}$ odhadnite

$J_1(2)$, a porovnajte svoj výsledok s výsledkom $J_1(2) = 0.576725$, ktorý je presný na šesť desatinných miest

V cvičeniach 8 až 10 použite integrálnu reprezentáciu

8. Odhadnite $J_2(2)$ použitím zloženého Simpsonovho pravidla s krokom dĺžky $h = \frac{\pi}{8}$, a porovnajte váš výsledok s $J_2(2) = 0.352834$, ktorá je presná na šesť desatinných miest.

9. Odhadnite $J_1(4)$ použitím zloženého Simpsonovho pravidla s krokom dĺžky $h = \frac{\pi}{10}$, a porovnajte váš výsledok s $J_1(4) = -0.066043$, ktorá je presná na šesť desatinných miest.

10. Odhadnite $J_3(4)$ použitím zloženého Simpsonovho pravidla s krokom dĺžky $h = \frac{\pi}{10}$, a porovnajte váš výsledok s $J_3(4) = 0.430171$, ktorá je presná na šesť desatinných miest.

11. Modifikonaná Besselova funkcia $I_0(x)$ má integrálnu reprezentáciu $I_0(x) = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} cosh(xsin\Theta)d\Theta$. Použite zložené Simpsonovo pravidlo s krokom dĺžky $h = \frac{\pi}{16}$, na odhad $I_0(3.5)$, a porovnajte váš výsledok s $I_0(3.5) = 7.378203$, ktorý je presný na šesť desatinných miest.

12. Modifikonaná Besselova funkcia $I_1(x)$ má integrálnu reprezentáciu $I_1(x) = \frac{2x}{\pi}\int_0^{\frac{\pi}{2}} cosh(xsin\Theta)\left(\cos\Theta\right)^2 d\Theta$. Použite zložené Simpsonovo pravidlo s krokom dĺžky $h = \frac{\pi}{16}$, na odhad $I_1(3)$, a porovnajte váš výsledok s $I_1(3) = 3.953370$, ktorý je presný na šesť desatinných miest.

V cvičeniach 13 a 16 použite $3, 5$, a 10 bodové Gaussove formuly na odhad daného integrálu a výsledky porovnajte s presnou hodnotou.

13. $I = \int_0^{\frac{3\pi}{2}} \cos x dx$. Presná hodnota je $I = -1$.

14. $I = \int_0^{\frac{\pi}{2}} e^{-x}\cos x dx$. Presná hodnota na šesť desatinných miest je $I = \frac{1}{2}[1 + exp(-\frac{\pi}{2})] = 0.603940$.

15. Použite 10-bodovú Gaussovu formulu na odhad hodnoty konvergennáho nevlastného integrálu $I = \int_0^{\frac{1}{2}} \frac{dx}{(1-4x^2)^{\frac{1}{2}}}$. Porovnajte výsledok s presnou hodnotou na šesť desatinných miest $I = \frac{\pi}{4} = 0.785398$.

16. Použite 10-bodovú Gaussovu formulu na odhad hodnoty konvergennáho nevlastného integrálu $I = \int_0^{\frac{\pi}{2}} \frac{\sqrt{x}dx}{\sin x}dx$. Porovnajte výsledok s presnou hodnotou na šesť desatinných miest $I = 2.753142$.

Chapter 5 NUMERICKÉ RIEŠENIE SÚSTAV LINEÁRNYCH ROVNíC.

Ukážeme dva prístupy k riešeniu sústav n nehomogénnych lineárnych rovníc s n neznámymi $x_1, x_2, ..., x_n$, a oba sú dôležité. Tieto metódy s rôznymi zjemneniami možno nájsť vo väčšine softvéroých balíkov lineárnej algebry. Prvá metóda zahŕňa postupnú elimináciu neznámych je priamy typ, v ktorom riešenie dostaneme po systematickom eliminovaní $n-1$ z n neznámych, aby sme našli x_n.Proces spätnej substitúcie sa použije na nájdenie zostávajúcich neznámych v opačnom poradí $x_{n-1}, x_{n-2}, ..., x_1$. Táto metóda môže byť použitá keď počet rovníc nie je rovný počtu neznámych, keď čo sa tiež ukáže automaticky, ak je systém inkonzistentný. Druhá metóda je prirodzene taká istá ako prvá s výnimkou cesty v ktorej detaily eliminačného procesu sú zaznamenané aby dovolili riešiť vhodne viac než jeden systém rovníc s tou istou maticou koeficientov. Aplikuje sa na systémy v ktorých sa počet rovníc rovná počtu neznámych. Prístup je taký, že sa pokúšame faktorizovať maticu koeficientov A v systéme $Ax = b$ na súčin $PA = LU$, kde L je dolná trojuholníková matica s $1-$ kami na svojej hlavnej diagonále, U je horná trojuholníková matica a P je permutačná matica, dôvod vysvetlíme neskôr. Metóda používa túto faktorizáciu na určenie riešenia - vektora x. Zlyhanie metódy ukáže, že táto faktorizácia naznačuje, že A je singulárna, teda jeden alebo viac jej riadkov sú lineárne závislé od ostatných riadkov.

Druhý typ prístupu je iteratívny, a aplikuje sa iba na sústavu n nehomogénnych rovníc s n neznámymi $x_1, x_2, ..., x_n$. Metódy štartujú s ľubovoľnou aproximáciou $x^{(0)}$ vektora riešenia x, a tento je iterovaný takým spôsobom, že to vedie ku postupným zlepšujúcim sa aproximáciám $x^{(1)}, ..., x^{(i)}, x^{(i+1)}$ riešenia x. The iteračný proces končí po N iteráciách, pokiaľ dve po sebe idúce postupné iterácie $x^{(N-1)}$ and $x^{(N)}$ vytvoria aproximácie $x_i^{(N-1)}$ až $x_i^{(N)}$ neznámej x_i, pre $i = 1, 2, ..., n$, sa líšia o menej ako predpísaná hodnota $\varepsilon > 0$, ktorú nazývame tolerancia. Konečná iterácia sa berie ako riešenie sústavy rovníc s vybranou toleranciou. Počet iterácií nutných na to, aby sme dostali takúto aproximáciu vektora riešení nie je určený, pretože závisí od štruktúry rovníc, zvolenej iteračnej schémy a tolerancie. Ako všetky metódy priameho typu sú v zmysle odvodené od štandardného Gaussovho eliminačného procesu, je potrebné popísať tento proces detailne. Neskôr ponúkneme modifikáciu procesu aby sme zaručili, že eliminačná procedúra neprekročí toleranciu v iteráciách nutnú aby hrubé (round-off) chyby boli minimalizované. Druhá priama metóda zachováva informácie obsiahnuté v Gaussovom eliminačnom procese a použije ich na odvodenie factorizácie $PA = LU$, po ktorej sa výsledok použije na riešenie sústavy $Ax = b$. Táto metóda je vhodná keď sa požadujú riešenia sústavy $Ax = b$ pre postupnosť nehomogénnych vektorov b zatiaľ čo koeficienty matice A zostávajú nezmenené. Toto sa môže stať napríklad pri analýze síl v štruktúre vďaka zmenám v zaťažení, kde matica A reprezentuje štruktúru, ktorá zostáva rovnaká, zatiaľ čo zaťaženie reprezentujúce vektorom b sa opakovane mení. Z množstva rôznych iteračných schém ktoré sú k dispozícii, popíšeme iba Jacobiho a Gauss–Seidel schémy. Tieto sú široko používané aj keď z iných dôvodov a sú aplikovateľné na systémy rovníc, ktoré majú vlastnosť nazývaná diagonálna dominancia. Iteračné metódy sa použijú, keď pracujeme s veľkými maticami, kde sa

často stáva, že matica obsahuje mnoho nulových prvkov, ktoré sa často vyskytujú na vedľajších diagonálach rovnobežných s hlavnou diagonálou matice A. Matice tohto typu nazývame sparse matrices (riedke matice), a vyskytujú sa pri riešení parciálnych diferenciálnych rovníc splinovou interpoláciou a v mnohých iných aplikáciách.

Gaussova eliminácia.

Predpokladáme, že sústava rovníc, ktorú máme riešiť má tvar

$$Ax = b, \tag{(25)}$$

Ak (25) prepíšeme explicitne máme

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{(26)}$$

Vieme, že (26), alebo (25), má jediné riešenie ak hodnosť matice A sa rovná hodnosti rozšírenej matice $[A|b]$ a obe sú rovné n. Ale potreba nájsť iný spôsob výpočtu x pochádza z faktu, že riešenia pomocou inverznej matice sú nepraktické pre n veľké, kvôli množstvu úloh pri výpočte A^{-1}. V oboch počítačovom aj ručnom počítaní nasledujúca plná matica v (26) je skrátená na rozšírenú maticu a výpočty sa potom realizujú na jej prvkoch. Rozšírená matica odpovedajúca (26) je

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{pmatrix} \tag{(27)}$$

V tomto skrátenom zápise koeficientov $x_1, x_2, ..., x_n$ v každej rovnici sú identifikované ich pozíciou v matici, teda koeficient x_1 je a_{12}, zatiaľ čo koeficient x_2 v $n-$tej rovnici je a_{n2}. Pretože jednotlivé rovnice môžu byť násobené k, a násobok rovnice môžme pričítať k inej rovnici, všetko bez zmeny riešenia, čo implikuje, že tieto isté operácie môžme urobiť na matici (27). Základný Gaussov eliminačný proces používa tieto vlastnosti. Prvý krok Gaussovho eliminačného procesu zahŕňa prepoklad $a_{11} \neq 0$, násobiac prvý riadok $\frac{a_{21}}{a_{11}}$, a odčítaním výsledku od druhého riadku, ak sa jeho prvý prvok stane nulou. Nasledujúci krok je násobiť prvý riadok $\frac{a_{31}}{a_{11}}$ a odčítať výsledok od tretieho riadku, v ktorom bude jeho prvý vstup nula. Opakovanie tohto procesu $n-1$ krát zakončuje prvú fázu-stage procesu, po ktorej všetky prvky pod a_{11} sú nuly, čo spôsobí, že (27) bude

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} & b_n^{(1)} \end{pmatrix}, \tag{(28)}$$

kde $a_{ij}^{(1)}$ a $b_i^{(1)}$ reprezentuje modifikované prvky a_{ij} a b_i po odčítaní násobku odpovedajúceho prvku v prvom riadku. Druhá etapa eliminačného procesu zahŕňa predpoklad, $a_{22}^{(1)} \neq 0$, odčítaním vhodných násobkov modifikovaného druhého riadku v (28) od $n-2$ riadkov pod dostaneme všetky prvky v stĺpci pod $a_{22}^{(1)}$ nulové. Pokračovnie tohto procesu predpokladajúc že žiadny prvok použitý na elimináciu prvkov pod ním nie je nula, vedie nakoniec k tomu, že všetky prvky pod vedúcimi diagonálnymi prvkami prvých n stĺpcov modifikovaného rozšíreného poľa bude nula, teda konečné rozšírené pole bude

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn}^{(n-1)} & b_n^{(n-1)} \end{pmatrix} \qquad ((29))$$

Riešenie potom nájdeme procesom nazvaným spätná substitúcia, ktorá štartuje z posledného riadku v (29), čo je ekvivalentné s rovnicou $a_{nn}^{(n-1)} x_n = b_n^{(n-1)}$, z ktorej plynie, že

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}}. \qquad ((30))$$

Druhý riadok odspodu v (29) je ekvivalentý s rovnicou

$$a_{n-1,n-1}^{(n-2)} x_{n-1j} + a_{n-1,n}^{(n-2)} x_n = b_{n-1}^{(n-2)}, \qquad ((31))$$

odkiaľ x_{n-1} nájdeme po substitúcii hodnoty x_n nájdenej v (30). Pokračovanie týmto spôsobom zaručí, že všetky prvky $x_1, x_2, ..., x_n$ riešenia x nájdeme v opačnom poradí $x_n, x_{n-1}, ..., x_1$. Prvky $a_{11}, a_{22}^{(1)}, ..., a_{nn}^{(n-1)}$ použité na redukciu koeficientov matice A do dolnej trojuholníkovej matice z prvých n stĺpcov (29) sa nazývajú pivoty Gaussovej eliminácie a riadok obsahujúci pivot sa nazýva pivotný riadok. Toto kompletizuje základný Gaussov eliminačný proces. Je jasné ak v $r-$ tom kroku procesu dostaneme riadok núl v modifikovanej matici koeficientov A, ale modifikovaný $r-$tý prvok v nehomogénnom vektore b je nenulový, sústava rovníc je nekompatibilná a žiadne riešenie neexistuje. Ak však v $r-$tom kroku eliminačného procesu dostaneme nulový riadok v modifikovanej matici koeficientov A, a modifikovaný $r-$tý prvok v nehomogénnom vektore je tiež nula, potom $r-$tá rovnica je lineárne závisla od prvých $r-1$ rovníc, teda riešenie nemôže byť jediné. Ťažkosť nastáva ak v ľubovoľnom kroku procesu pivot v $m-$tej pozícii na hlavnej diagonále modifikovanej matice A bude nula, podobne ako sa stane na štarte ak $a_{11} = 0$. Ak by sa to stalo, ťažkosť možno prekonať zámenou poradia ridkov aby sme dostali nenulový prvok na pozíciu pivota. Chyby v eliminačnom procese môžu nastať ak veľmi malé pivoty sa použijú na redukciu nulových prvkov v stĺpcoch pod ak sú tieto značne väčšie, čomu treba zabrániť. Ak poradie rovníc musí byť zmenené bez zmeny riešenia, táto nevýhoda sa dá odstrániť nasledovne: ak v $m-$tom kroku z ostatných riadkov $m-$tý na $n-$tý je vybraný a obsahuje jeden z prvkov s väčšou hodnotou sa vyberie ako prvok s najväčšou veľkosťou v $m-$tom stĺpci.Tento riadok sa potom posunie nahor aby vytvoril nový $m-$tý riadok po ktorom eliminačný proces pokračuje ako predtým. Tento proces nazývame Gaussova eliminácia s parciálnym pivotingom, a je štandardnou súčasťou softvérov. Ľahko vidieť, že

podobnú metódu možno použiť, ak sa počet rovníc nerovná počtu neznámych. Potom nájdeme modifikovanú rozšírenú maticu a zistíme , či sústava nemá riešenie, má jediné riešenie, alebo viac riešení závisiacich od parametrov. Aj keď $\det A$ v G.E.M. nevyžadujeme pretože proces redukuje originálne koeficienty matice A na účinný spôsob G.E.M. $\det A$ dolnej trojuholníkovej matice

$$\det A = a_{11}a_{22}^{(1)}a_{33}^{(2)}...a_{nn}^{(n-1)}, \qquad ((32))$$

a je to táto metóda, ktorá sa používa v softwarových programoch na hľadanie $\det A$, aby sme sa vyhli mnohonásobnému čas žerúcemu násobeniu pri výpočte kofaktorov.

Example 18 *Riešme nasledujúcu sústavu rovníc Gaussovou eliminániou:*

$$\begin{array}{rrrrl} 2x_1 & -2x_2 & +3x_3 & +4x_4 & = -18 \\ 4x_1 & +x_2 & -x_3 & +2x_4 & = -11 \\ x_1 & -x_2 & -x_3 & +5x_4 & = -26 \\ 2x_1 & -3x_2 & +2x_3 & -x_4 & = -3 \end{array} .$$

Použite (32) na výpočet determinantu koeficientov matice A.

Solution 19 *Uvažujeme pole (maticu)*

$$\begin{pmatrix} 2 & -2 & 3 & 4 & -18 \\ 4 & 1 & -1 & 2 & -11 \\ 1 & -1 & -1 & 5 & -26 \\ 2 & -3 & 2 & -1 & -3 \end{pmatrix}$$

Pomocou $a_{11} = 2$ sa zbavíme prvkov v prvom stĺpci pod.

$$\begin{pmatrix} 2 & -2 & 3 & 4 & -18 \\ 0 & 5 & -7 & -6 & 25 \\ 0 & 0 & -\frac{5}{2} & 3 & -17 \\ 0 & -1 & -1 & -5 & 15 \end{pmatrix}$$

Podobne s pivotom 5:

$$\begin{pmatrix} 2 & -2 & 3 & 4 & -18 \\ 0 & 5 & -7 & -6 & 25 \\ 0 & 0 & -\frac{5}{2} & 3 & -17 \\ 0 & 0 & -\frac{12}{5} & -\frac{31}{5} & 20 \end{pmatrix}$$

Pokračujeme s -$\frac{-5}{2}$ až dostaneme:

$$\begin{pmatrix} 2 & -2 & 3 & 4 & -18 \\ 0 & 5 & -7 & -6 & 25 \\ 0 & 0 & -\frac{5}{2} & 3 & -17 \\ 0 & 0 & 0 & -\frac{227}{25} & \frac{908}{25} \end{pmatrix}$$

Spätnou substitúcoiu dostaneme riešenie: $-\frac{227}{25}x_4 = \frac{908}{25}$, teda $x_4 = -4$, podobne z $-\frac{5}{2}x_3 + 3x_4 = -17$, máme $x_3 = 2$. Pokračujúc podobne dostaneme $x_2 = 3$ a $x_1 = -1$. Poznamenajme, že žiadny pivot nebol dosť malý a parciálny pivoting nebol potrebný. $\det A$ *dostaneme priamo z (32):* $\det A = 2 \cdot 5 \cdot \left(-\frac{5}{2}\right) \cdot \left(-\frac{277}{25}\right) = 277.$

LU faktorizácia.

Nech A je nesingulárna matica typu $n \times n$ v sústave $Ax = b$, ktorá sa dá faktorizovať ako súčin $A = LU$, kde L je dolná trojuholníková matica typu $n \times n$ s $1-$kami na jej hlavnej diagonále a U je horná trojuholníková matica typu $n \times n$. Metóda riešenia sústavy rovníc $Ax = b$ sa redukuje na hľadanie stĺpcového vektora y takého, že je riešením $Ly = b$, a potom určenia x zo sústavy rovníc $Ux = y$. Výhodou tohto prístupu je, že ak už raz máme L a U, prvky vektora y nájdeme spätnou substitúciou, po ktorej prvky vektora x vypočítame opäť spätnou substitúciou. Ako sme už poznamenali, tento prístup je veľmi účinný, keď sústavu $Ax = b$ riešime opakovane s tou istou maticou A, ale s rôznymi nehomogénnymi vektormi b. Je to preto, lebo L a U zostanú nezmenené, teda aj vektor riešenia x môžeme nájsť iba použitím násobenia vektora b, a známej faktorizácie A. Poznamenajme, že bez riadkových permutácií nemusí byť možné faktorizovať nesingulárnu maticu. Všetky informácie nutné pre faktorizáciu A na súčin $A = LU$ sú obsiahnuté v Gaussovej eliminácii, teda najpriamejšia forma LU faktorizácie, v ktorej použijeme parciálny pivoting nie je nutná a ukážeme ju na predošlom príklade. Budeme faktorizovať maticu A z predošlého príkladu, a potom výsledok použijeme na riešenie sústavy rovníc v tomto príklade.

Keď sme použili prvú etapu Gaussovej eliminácie na matici A v príklade 2 krát riadok 1 bol odčítaný od riadku 2, $\frac{1}{2}$ riadku 1 bola odčítaná od riadku 3, a 1 krát riadok 1 bol odčítaný od riadku 4, čo spôsobilo zmenu matice

$$A = \begin{pmatrix} 2 & -2 & 3 & 4 \\ 4 & 1 & -1 & 2 \\ 1 & -1 & -1 & 5 \\ 2 & -3 & 2 & -1 \end{pmatrix} \text{ na maticu } A_1 = \begin{pmatrix} 2 & -2 & 3 & 4 \\ 0 & 5 & -7 & -6 \\ 0 & 0 & -\frac{5}{2} & 3 \\ 0 & -1 & -1 & -5 \end{pmatrix}.$$

Ak reprezentujeme elementárne riadkové operácie pomocou prenásobení A maticou M_1, môžme to zapísať $M_1 A = A_1$, kde

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -\frac{1}{2} & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.$$

Aplikáciou druhej etapy GEM na A_1, $\frac{-1}{5}$ krát riadok 2 bol odpočítaný od riadku 4, čo spôsobilo maticu A_2, tvaru

$$A_2 = \begin{pmatrix} 2 & -2 & 3 & 4 \\ 0 & 5 & -7 & -6 \\ 0 & 0 & -\frac{5}{2} & 3 \\ 0 & 0 & -\frac{12}{5} & -\frac{31}{5} \end{pmatrix},$$

čo vo vyjadrení násobenia matíc dáva $M_2 A_1 = A_2$, alebo $M_2 M_1 A = A_2$, kde

$$M_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{5} & 0 & 1 \end{pmatrix}.$$

Nakoniec ak v poslednej etape GEM was applied to matrix A_2, $\frac{24}{25}$ krát riadok 3 bol odčítaný od riadku 4 dostali sme hornú trojuholníkovú maticu

$$A_3 = \begin{pmatrix} 2 & -2 & 3 & 4 \\ 0 & 5 & -7 & -6 \\ 0 & 0 & -\frac{5}{2} & 3 \\ 0 & 0 & 0 & -\frac{227}{25} \end{pmatrix},$$

čo v reči násobenia matíc zapíšeme ako $M_3A_2 = A_3$, alebo $M_3M_2M_1A = A_3$, kde

$$M_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{24}{25} & 1 \end{pmatrix}.$$

Tak máme $A_3 = U$ je horná trojuholníková matica a ukázali sme, že $M_3M_2M_1A = U$, s

$$U = \begin{pmatrix} 2 & -2 & 3 & 4 \\ 0 & 5 & -7 & -6 \\ 0 & 0 & -\frac{5}{2} & 3 \\ 0 & 0 & 0 & -\frac{227}{25} \end{pmatrix},$$

a teda $A = M_1^{-1}M_2^{-1}M_3^{-1}U$. Dokázali sme faktorizovať A a ak ukážeme, že $M_1^{-1}M_2^{-1}M_3^{-1}$ je dolná trojuholníková matica požadovaného typu. Aby sme to dosiahli poznamenajme, že špeciálna štruktúra matíc M_i, for $i = 1, 2, 3$ je taká, že z definície inverznej matice pomocou jej kofaktorov, inverznú maticu M_i^{-1} dostaneme priamo z matice M_i zmenou znamienka prvkov v jej $i-$ tom stĺpci, ktoré ležia pod prvkom 1, tak bez ďalších výpočtov máme

$$M_1^{-1}M_2^{-1}M_3^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\frac{1}{5} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{24}{25} & 1 \end{pmatrix}.$$

Štruktúra týchto matíc dovoľuje zapísať ich súčin ihneď, pretože $i-$ty stĺpec súčinu matíc je jednoducho $i-$ty stĺpec matice M_i, teda

$$M_1^{-1}M_2^{-1}M_3^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 \\ 1 & -\frac{1}{5} & \frac{24}{25} & 1 \end{pmatrix}.$$

Je to dolná trojuholníková matica požadovaného tvaru, teda

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 \\ 1 & -\frac{1}{5} & \frac{24}{25} & 1 \end{pmatrix},$$

a faktorizovaný tvar A je

$$A = LU = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 \\ 1 & -\frac{1}{5} & \frac{24}{25} & 1 \end{pmatrix} \begin{pmatrix} 2 & -2 & 3 & 4 \\ 0 & 5 & -7 & -6 \\ 0 & 0 & -\frac{5}{2} & 3 \\ 0 & 0 & 0 & -\frac{227}{25} \end{pmatrix}.$$

Použiť L a U na riešnie sústavy z príkladu, musíme najprv riešiť sústavu $Ly = b$, kde $b = [-18, -11, -26, -3]^T$. Je to sústava

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 \\ 1 & -\frac{1}{5} & \frac{24}{25} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} -18 \\ -11 \\ -26 \\ -3 \end{pmatrix},$$

from which we see that $y_1 = -18$, and forward substitution then shows $y_2 = 25, y_3 = -17$ a $y_4 = \frac{908}{25}$. Prvky x_1, x_2, x_3, x_4 riešenia x teraz dostaneme riešením sústavy $Ux = y$:

$$\begin{pmatrix} 2 & -2 & 3 & 4 \\ 0 & 5 & -7 & -6 \\ 0 & 0 & -\frac{5}{2} & 3 \\ 0 & 0 & 0 & -\frac{227}{25} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -18 \\ 25 \\ -17 \\ \frac{908}{25} \end{pmatrix}.$$

Odkiaľ $x_4 = -4$, a spätnou substitúciou dostaneme $x_3 = 2, x_2 = -3, x_1 = -1$, a sústava je vyriešená.

LU factorizačný algoritmus.
Faktorizáciu nesingulárnej matice A typu $n \times n$ na súčin $A = LU$, kde L je dolná troluholníková matica s $1-$kami na hlavnej diagonále a U je horná trojuholníková matica možno urobiť nasledovne:

1. Maticu U dostaneme aplikáciou GEM na riadky A a tým ju zredukujeme na hornú trojuholníkovú maticu.

2. V $i-$tej etape GEM v kroku 1 a v $i-$tom stĺpci nech m_{ij} je násobok $i-$teho prvku ktorý má byť odčítaný od $j-$teho prvku, aby sa redukoval $j-$ty prvok na nulu. Potom matica L je daná

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ m_{21} & 1 & 0 & 0 & \cdots & 0 & 0 \\ m_{31} & m_{32} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1 & \vdots \\ m_{n1} & m_{n2} & m_{n3} & m_{n4} & \cdots & m_{n,n-1} & 1 \end{pmatrix}$$

Example 20 *Aplikujte LU faktorizačný algoritmus na určenie matice L v predošlom príklade.*

Solution 21 *Skúmaním GEM procesu v predošlom príklade ukazuje v prvom kroku $m_{21} = 2, m_{31} = \frac{1}{2}, m_{41} = 1$, a v druhom kroku $m_{32} = 0, m_{42} = -\frac{1}{5}$, zatiaľ čo v poslednom kroku $m_{43} = \frac{24}{25}$, tak z algoritmu*

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 \\ 1 & -\frac{1}{5} & \frac{24}{25} & 1 \end{pmatrix}.$$

Jacobiho iteračný proces.

Odvodiť Jacobiho iteračný proces, jednotlivé rovnice v (26) preformulujeme do tvaru

$$\begin{aligned} x_1 &= \frac{(b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n)}{a_{11}} \\ x_2 &= \frac{(b_2 - a_{21}x_1 - a_{23}x_3 - \cdots - a_{2n}x_n)}{a_{22}} \\ &\vdots \\ x_n &= \frac{(b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{nn-1}x_{n-1})}{a_{nn}} \end{aligned} \qquad ((33))$$

Jacobiho iteračný proces dostaneme z tohto vyjadrenia ak definujeme $r-$tú aproximáciu riešenia, ktorú označíme $x_1^{(r)}, x_2^{(r)}, ..., x_n^{(r)}$, pomocou $(r-1)-$vej aproximácie $x_1^{(r-1)}, x_2^{(r-1)}, ..., x_n^{(r-1)}$, pomocou rovníc Jacobiho iteračnej metódy

$$\begin{aligned} x_1^{(r)} &= \frac{(b_1 - a_{12}x_2^{(r-1)} - a_{13}x_3^{(r-1)} - \cdots - a_{1n}x_n^{(r-1)})}{a_{11}} \\ x_2^{(r)} &= \frac{(b_2 - a_{21}x_1^{(r-1)} - a_{23}x_3^{(r-1)} - \cdots - a_{2n}x_n^{(r-1)})}{a_{22}} \\ &\vdots \\ x_n^{(r)} &= \frac{(b_n - a_{n1}x_1^{(r-1)} - a_{n2}x_2^{(r-1)} - \cdots - a_{nn-1}x_{n-1}^{(r-1)})}{a_{nn}}. \end{aligned} \qquad ((34))$$

Iterácia štartuje nejakým začiatočným výberom $x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}$, typicky $x_1^{(0)} = 1, x_2^{(0)} = 1, ..., x_n^{(0)} = 1$. Iteračný proces pokračuje pokiaľ pre nejaké r diferencia medzi odpvedajúcimi prvkami $(r-1)-$vej a $r-$tej iterácie: $\left|x_i^{(r)} - x_i^{(r-1)}\right|$, pre $i = 1, 2, ..., n$ je menšia než dopredu stanovená tolerancia $\varepsilon > 0$, teda

$$\left|x_i^{(r)} - x_i^{(r-1)}\right| < \varepsilon, \quad \text{pre} \;\; i = 1, 2, ..., n. \qquad ((35))$$

Toto je najjednoduchšia z mnohých podmienok konvergencie iteračného procesu. Hodnoty $x_1^{(r)}, x_2^{(r)}, ..., x_n^{(r)}$ získané z $r-$tej iterácie v ktorej sú splnené podmienky (35) prvý krát sú považované za riešenie $x_1, x_2, ..., x_n$, v tolerancii ε. Treba poznamenať, že Jacobiho iteračný proces je iteračný proces pevného bodu pre sústavy lineárnych rovníc. Aj keď to nebudeme dokazovať, postačujúcou podmienkou konvergencie Jacobiho iteračného procesu pre ľubovoľný výber $x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}$ je, že sústava (26) je diagonálne dominantá, čo znamená, že v každom riadku koeficienty diagonálnej dominancie matice A, t.j. veľkosť prvkov ležiacich na hlavnej

diagonále presahuje súčet veľkostí ostatných prvkov v riadku. Teda matica A bude diagonálne dominantná ak

$$|a_{ii}| > |a_{i1}| + |a_{i2}| + \cdots + |a_{ii-1}| + |a_{ii+1}| + \cdots + |a_{in}|\,, \text{ pre } i = 1, 2, ..., n \quad ((36))$$

Overenie rovníc v (34) ukazuje, že Jacobiho metóda zlyhá, ak sa použijú bežné (improved) aproximácie ako sú generované. Toto je vidieť v druhej rovnici, kde lepší odhad $x_1^{(r)}$ nájdený z prvej rovnice môže byť použitý namiesto odhadu $x_1^{(r-1)}$. Postupujúc takýmto spôsobom ak v každej rovnici vždy použijeme bežne dostupný odhad, povedie ku Gauss–Seidel iteračnému procesu definovanému Gauss–Seidel iteračnou metódou

$$\begin{aligned}
x_1^{(r)} &= \frac{b_1 - a_{12}x_2^{(r-1)} - a_{13}x_3^{(r-1)} - \cdots - a_{1n}x_n^{(r-1)}}{a_{11}} \\
x_2^{(r)} &= \frac{b_2 - a_{21}x_1^{(r)} - a_{23}x_3^{(r-1)} - \cdots - a_{2n}x_n^{(r-1)}}{a_{22}} \qquad ((37)) \\
x_3^{(r)} &= \frac{b_3 - a_{31}x_1^{(r)} - a_{32}x_2^{(r)} - \cdots - a_{3n}x_n^{(r-1)}}{a_{33}} \\
&\vdots \\
x_n^{(r)} &= \frac{b_n - a_{n1}x_1^{(r)} - a_{n2}x_2^{(r)} - \cdots - a_{nn-1}x_{n-1}^{(r)}}{a_{nn}}.
\end{aligned}$$

Postačujúca podmienka pre konvergenciu Gauss–Seidel procesu je taká istá ako pre Jacobiho proces, a to, že A je diagonálne dominantá. Ostatné podmienky konvergencie iteračného procesu môžme odvodiť v reči magnitúdy najväčšej vlastnej hodnoty A, ktorú nazývame spektrálny polomer, ale spektrálny polomer ako túto vlastnú hodnotu je ťažké vypočítať, keď je počet rovníc n veľký, také výsledky majú najmä teoretickú dôležitosť. Ak iteračný proces diverguje, postupné iterácie zvyčajne menia znamienko a ich veľkosť rastie bez ohraničenia. V softwarových programoch býva kontrola ohľadne chovania postupných iterácií a ak sa zistí divergencia, počítač o tomto podá odkaz a ukončí výpočet.

Example 22 *Použite Gaussov–Seidelov iteračný proces a nájdite riešenie nasledujúcej sústavy rovníc:*

$$\begin{aligned}
1.2x_1 \quad &+4.4x_2 \quad -1.9x_3 \quad = -4.2 \\
5.1x_1 \quad &-1.3x_2 \quad +2.4x_3 \quad = 2.7 \\
-2.6x_1 \quad &+1.7x_2 \quad -6.3x_3 \quad = 9.6
\end{aligned}$$

Solution 23 *Aplikácia testu diagonálnej dominancie v (36) ukazuje, že iba tretia rovnica vyhovuje podmienkam, pretože* $|-6.3| > |-2.6| + |1.7|$, *ale* $|1.2| < |4.4| + |-1.9|$ *a* $|-1.3| < |5.1| + |2.4|$. *Avšak ak v prvých dvoch rovniciach are interchanged sústava sa stane diagonálne dominantou, teda ak položíme v Gauss–Seidel iteračnom procese v tomto prípade rovnice musia byť použité v poradí*

$$\begin{aligned} 5.1x_1 \quad -1.3x_2 \quad +2.4x_3 \quad &= 2.7 \\ 1.2x_1 \quad +4.4x_2 \quad -1.9x_3 \quad &= -4.2 \\ -2.6x_1 \quad +1.7x_2 \quad -6.3x_3 \quad &= 9.6 \end{aligned}$$

Z (37) Gauss–Seidelov iteračný proces pre tento systém rovníc bude:

$$x_1^{(r)} = \frac{2.7 + 1.3x_2^{(r-1)} - 2.4x_3^{(r-1)}}{5.1}$$

$$x_2^{(r)} = \frac{-4.2 - 1.2x_1^{(r)} + 1.9x_3^{(r-1)}}{4.4}$$

$$x_3^{(r)} = \frac{9.6 + 2.6x_1^{(r)} - 1.7x_2^{(r)}}{-6.3}$$

Výsledky štartujúce z iterácií $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 1$ *sú v nasledujúcej tabuľke a hodnoty získané v 10 iterácii môžu byť porovnané s riešením získaným G.E.M.:*$x_1 = 1.162946, x_2 = -2.418817, x_3 = -2.656452$.

Počet iterácií

	0	1	2	3	4	5
x_1	1	0.313726	1.229617	1.219913	1.175631	1.162857
x_2	1	−0.608289	−2.074693	−2.406137	−2.430951	−2.422740
x_3	1	−1.817425	−2.591108	−2.676541	−2.664962	−2.657887

	6	7	8	9	10
x_1	1.162621	1.162815	1.162924	1.162946	1.162947
x_2	−2.419348	−2.418785	−2.418784	−2.418809	−2.418816
x_3	−2.656461	−2.656389	−2.656434	−2.656450	−2.656452

Tieto výsledky ukazujú konvergenciu iteračnej metódy získanej z diagonálne dominantnej schémy oproti priamej metóde. Ak miesto iteračnej schémy vyjdeme z originálnej sústavy rovníc bez novej reformulácie, potom by sme dostali nediagonálne dominantný systém, ktorý vedie na divergentný systém.

$$x_1^{(r)} = \frac{4.2 - 4.4x_2^{(r-1)} + 1.9x_3^{(r-1)}}{1.2}$$

$$x_2^{(r)} = \frac{-2.7 - 5.1x_1^{(r-1)} - 2.4x_3^{(r-1)}}{-1.3}$$

$$x_3^{(r)} = \frac{9.6 + 2.6_1^{(r-1)} - 1.7x_2^{(r-1)}}{-6.3}$$

Použitím tejto schémy a štartovacích bodov ako predtým $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 1$, *dostaneme výsledok:*

$x_1^{(1)} = -5.58333, x_2^{(1)} = -22.13462, x_3^{(1)} = -5.19241$

$x_1^{(2)} = 69.43894, x_2^{(2)} = 260.75140, x_3^{(2)} = 40.18034$

čo jasne demonštruje divergenciu nediagonálne dominantnej schémy .

Musíme povedať ako sa tieto dve iteračné metódy používajú. Gauss–Seidel metóda sa používa hlavne v computerových výpočtoch ako predchádzajúca podmienka (predpoklad) pre viac rozvinuté schémy, kde jej použitie bežných aproximácií v každom kroku vyžaduje iba polovicu pamäti ako ako Jacobiho metóda. Jacobiho schémy sa používajú extenzívne na budovanie blokov v komplikovanejších a neúplných iteračných procedúrach, ako predpokladaný conjugovaný gradient a multigrid metódy.

Zhrnutie: V rôznych príkladoch sme videli, že LU faktorizácia $n \times n$ matice A je možná, iba ak $\det A = 0$. Dve zásadne odlišné typy metód boli odvodené na riešenie sústav nehomogénnych lineárnych rovníc, prvá priama, druhá iteračná. Dve priame metódy boli Gaussova eliminácia a z nej odvodená LU faktorizácia. Ostatné metódy vrátane iteračných štartovali z ľubovoľných začiatočných aproximácií a konvergovali k požadovanému riešeniu v rámci predpísanej tolerancie, za predpokladu diagonálnej dominancie.

Cvičenia.

V cvičeniach 1. - 4. a) riešte sústavu rovníc použitím Gaussovej eliminačnej metódy, b) porovnajte výsledok s a) s tou, ktorú dostanete riešením sústavy Gaussovou -Seidelovou metódou so štartovacou hodnotou $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 1$ a urobte 10 iterácií.

1. $4.7x_1 + 1.3x_2 - 1.6x_3 = 1.3$
   $x_1 - 4.1x_2 + 1.1x_3 = 4.6$
   $2.1x_1 + 1.4x_2 + 6.2x_3 = 5.2.$

2. $1.7x_1 - 4.6x_2 - 1.2x_3 = 3.4$
   $-3.1x_1 + 2.3x_2 + 7.2x_3 = 2.7$
   $3.2x_1 + 1.2x_2 + 1.4x_3 = -4.2.$

3. $2.1x_1 + 6.5x_2 - 3.1x_3 = -6.4$
   $-5.2x_1 + 2.1x_2 - 1.5x_3 = 3.7$
   $1.8x_1 - 2.9x_2 + 6.2x_3 = -4.2.$

4. $6.2x_1 - 2.2x_2 + 3.1x_3 = -2.6$
   $-1.6x_1 + 1.9x_2 + 8.4x_3 = -2.6$
   $2.3x_1 - 8.4x_2 + 3.2x_3 = 6.5.$

5. Reálna $n \times n$ symetrická matica H_n s prvkami $h_{ij} = \frac{1}{(i+j-1)}$ v $i-$tom riadku a $j-$teho stĺpca sa nazýva Hilbertova matica, a jeho determinant sa rýchlo stáva veľmi malým ako n rastie. Matice tohto typu sa nazývajú zle podmienené a keď sa zle podmienené matice stanú maticami koeficientov sústavy lineárnych rovníc, nastávajú veľké chyby aj keď sa výpočty vykonávajú s použitím veľkej presnosti. Rozvoj veľmi malých determinantov Hilbertovej matice možno vidieť napríklad ak $n = 4$, pretože

   $$H_4 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix} \qquad \text{a det } H_4 = \frac{1}{6{,}048{,}000.}.$$

   Ak zlomky v matici neaproximujeme, presné riešenie sústavy rovníc

   $$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

   bude $x_1 = -64, x_2 = 900, x_3 = -2520, x_4 = 1820$. Typicky zle podmienená matica sa vyskytuje v prípade aproximácie metódou najmenších štvorcov a ortogonalizáciou. Demonštrujte chyby, ktoré sa vyskytnú, keď sa použije Gaussova eliminácia na riešenie sústavy rovníc a výpočty sa zaokrúhľujú na päť desatinných miest. Použite G.E.M. na výpočet det H_4 pracujúc na päť desatinných miest a porovnajte s hodnotou správneho výsledku.

6. Použite Jacobiho a Gauss–Seidelove iterácie na riešenie sústavy $-4.2x_1 + 1.1x_2 - 2.1x_3 = 1.4$

   $3.6x_1 + 9.2x_2 - 3.1x_3 = -3.2$

   $1.4x_1 + 2.9x_2 - 6.4x_3 = -1.2,$

   štartujúc zo začiatočnej iterácie $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$ a vykonajte šesť iterácií. Porovnajte výsledky s presným riešením $x_1 = -0.39101, x_2 = -0.18938, x_3 = 0.01615$. Odvoďte iteračnú schému keď rovnice preskupíme do nediagonálne dominantnej formy a použitím začiatočných hodnôt $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$ urobte tri iterácie na demonštrovanie divergentnosti schémy.

   V cvičeniach 7 - 12 použite LU faktorizáciu na riešenie sústavy rovníc $Ax = b$ pre dané matice A a b.

7. $A = \begin{pmatrix} -4 & 1 & -1 \\ 12 & -1 & 5 \\ -12 & 5 & -4 \end{pmatrix}, b = \begin{pmatrix} 3 \\ -2 \\ 2 \end{pmatrix}.$

8. $A = \begin{pmatrix} -1 & 2 & 3 \\ -5 & 7 & 16 \\ 2 & -10 & -2 \end{pmatrix}, b = \begin{pmatrix} -5 \\ 2 \\ 6 \end{pmatrix}.$

9. $A = \begin{pmatrix} 4 & -1 & -1 \\ -16 & 6 & 1 \\ -4 & 7 & -9 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 6 \\ -7 \end{pmatrix}.$

10. $A = \begin{pmatrix} -5 & -2 & 0 \\ -15 & -9 & 2 \\ 0 & -6 & 8 \end{pmatrix}, b = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}.$

11. $A = \begin{pmatrix} 2 & 1 & 0 & 2 \\ -1 & 0 & 1 & 0 \\ 4 & \frac{3}{2} & 2 & 3 \\ -2 & 0 & 8 & 1 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}.$

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

Vlastné hodnoty a vlastné vektory.

Vlastná hodnota a vlastný vektor súvisiaci s $n \times n$ maticou A bol definovaný ako číslo λ spĺňajúce maticovú rovnicu

$$Ax = \lambda x, \tag{(38)}$$

a odpovedajúci $n \times 1$ vektor x sme definovali ako zodpovedajúci vlastný vektor. Priamo z (38) plynie, že vlastný vektor x matice A odpovedajúci vlastnej hodnote λ možno násobiť (skalárom) t.j. nenulovým číslom k a stále zostane vlastným vektorom, pretože $A(kx) = \lambda(kx)$ je ekvivalentné s $kAx = k\lambda x$, a škrtnutie skalára k redukuje posledný výsledok na (38). Keď definujeme vlastné hodnoty a vlastné vektory, výsledok (38) sa dá napísať ako homogénny systém $(A - \lambda I)x = 0$, a vlastné hodnoty nájdeme ako $\det(A - \lambda I) = 0$, čo vedie k polynómu $P(\lambda)$ stupňa n tvaru $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n$, ktorý nazývame charakteristický polynóm A. Ak nájdeme nulové body $P(\lambda)$ sú to vlastné hodnoty $\lambda_1, \lambda_2, ..., \lambda_n$ matice A, a pridružené vlastné vektory $x_1, x_2, ..., x_n$ dostaneme riešením maticovej rovnice

$$Ax_i = \lambda_i x_i, \quad \text{pre } i = 1, 2, ..., n. \tag{(39)}$$

Tento teoretický prístup je výhodný ak $n \leq 3$, pretože vtedy nulové body $P(\lambda)$ možno určiť analyticky. Vo všetkých ostatných prípadoch je hľadanie nulových bodov ťažkou úlohou a aj keď sú známe presne, veľké chyby môžeme urobiť, ak ich použijeme v (39) na výpočet príslušných vlastných vektorov. Výpočtovo efektívne metódy na výpočet vlastných hodnôt a vlastných vektorov sú k dispozícii v balíkoch počítačových algebier, ktoré neumožňujú najskôr riešiť charakteristickú rovnicu pre vlastné hodnoty. Tieto su capable na hľadanie reálnych a komplexných vlasných hodnôt vrátane opakujúcich sa vlastných hodnôt a odpovedajúcich vektorov. Kvôli tomu jediná metóda, ktorú to popíšeme bude mocninová metóda, ktorá je ľahko aplikovateľná a jej odvodenie je priame. Ale nie je to metóda prakticky použiteľná okrem niektorých špeciálnych situácií. Odvodenie vyžaduje všetky aby vlastné hodnoty A boli usporiadané podľa veľkosti

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \leq |\lambda_n|. \tag{(40)}$$

Ak máme vlastné hodnoty usporiadané, λ_1 s najväčšou vlastnou hodnotou sa nazýva dominantná vlastná hodnota matice A, a ostatné vlastné hodnoty $\lambda_2, \lambda_3, ..., \lambda_n$ potom nazývame subdominantné vlastné hodnoty. Ľubovoľný n prvkový stĺpcový vektor It was seen in Chapter4 that an arbitrary n element column vector v_0 can always be expressed as the linear combination of eigenvectors $x_1, x_2, ..., x_n$,

$$v_0 = c_1x_1 + c_2x_2 + \cdots + c_nx_n, \tag{(41)}$$

pre nejaký vhodný výber konštánt $c_1, c_2, ..., c_n$. Mocninová metóda pre súčasné určenie vlastných hodnôt a vlastných vektorov A iteračnou metódou a zahŕňa substitúciu $v_r = A^r v_0$, násobením (41) A^r, a použitím výsledkov (39) a (40). Pre $r = 0, 1, 2, ...$,máme

$$\begin{aligned} v_r &= A^r(c_1x_1 + c_2x_2 + \cdots + c_nx_n) = \\ &= c_1\lambda_1^r x_1 + c_2\lambda_2^r x_2 + \cdots + c_n\lambda_n^r x_n = \\ &= \lambda_1^r \left\{ c_1x_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^r x_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^r x_n \right\} \end{aligned} \tag{(42)}$$

Usporiadanie vlastných hodnôt v (40) spôsobuje, že faktory $\left(\frac{\lambda_2}{\lambda_1}\right)^r, \left(\frac{\lambda_3}{\lambda_1}\right)^r, \ldots, \left(\frac{\lambda_n}{\lambda_1}\right)^r \to 0$ pre rastúce r, teda ak predpokladáme, že $c_1 \neq 0$, pre dostatočne veľké r rovnicu (42) možno aproximovať

$$x_r \approx \lambda_1^r c_1 x_1. \qquad ((43))$$

Predpoklad, že $c_1 \neq 0$ nie je reštriktívny, pretože ak je pravdivý roundoff možno očakávať zavedenie komponenty v smere x_1, teda aj keď konvergencia sa bude oneskorovať, prakticky nastane. Výsledok (43) ukayuje, že ak je r veľké, v_r možno vziať tak, aby bol úmerný odpovedajúceho prvku vo v_r a v_{r-1} aproximuje dominantú vlastnú hodnotu λ_1. Ak implementujeme mocninovú metódu prvky v_r môžu byť veľmi veľké, alebo veľmi malé, tak udržať aby rozsah exponentov v stroji nebol prekročený, možno vlastný vektor nechať škálovaný a stále zostane vlastným vektorom. Mnoho normalizácií je možné použiť, ale najvýhodnejšia dovoľuje získať $\widetilde{v_{r-1}}$ z v_{r-1} delením každého prvku v_{r-1} hodnotou α_{r-1}, kde α_{r-1} je prvok s najväčšou magnitúdou. Ako výsledok takejto normalizácie $v_{r-1} = \alpha_{r-1}\widetilde{v_{r-1}}$, a prvky vo v_{r-1} s najväčšou magnitúdou budú 1. Iteračná rovnica $v_r = Av_{r-1}$ musí byť zamenená za $v_r = A\widetilde{v_{r-1}}$ pre $r = 1, 2, ...$, alebo ekvivalentne za

$$v_{r+1} = A\widetilde{v_r} \quad \text{pre} \quad r = 0, 1, ... \qquad ((44))$$

Substitúciou $v_{r+1} = \alpha_{r+1}\widetilde{v_{r+1}}$ v predošlom vzťahu dáva $A\widetilde{v_r} = \alpha_{r+1}\widetilde{v_{r+1}}$, teda ak r bude veľké a $\widetilde{v_r} \to \widetilde{v_{r+1}},$, potom $\alpha_{r+1} \to \lambda_1$ a $\widetilde{v_{r+1}} \to \widetilde{x_1}$, normalizovaný vlastný vektor odpovedajúci λ_1. Iteračný proces v (44) môže štartovať s každým konštantným vektorom $v_0 = [v_1, v_2, ..., v_n]^T$, často sa berie $v_0 = [1, 1, ..., 1]^T$. Rýchlosť konvergencie iterácií sa zrýchľuje ak $|\lambda_1| \gg |\lambda_2|$, ale konvergencia je veľmi pomalá ak sú $|\lambda_1|$ a $|\lambda_2|$ blízke.

Example 24 *Použitím mocninovej metódy nájdite dominantnú vlastnú hodnotu λ_1 a normalizovaný vlastný vektor x_1 ak*

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

Solution 25 *Pretože je A symetrická matica, jej vlastné hodnoty budú reálne, teda je vhodné použiť mocninovú metódu na určenie vlastných hodnút a vlastných vektorov. Aby sme určili dominantnú vlastnú hodnotu a jej odpovedajúci vlastný vektor iteračný proces $v_r = A\widetilde{v_{r-1}}$ bude štartovať $v_0 = [1, 1, 1, 1]^T$, a v nasledujúcej tabuľke i-ty prvok v_r označíme $v_r^{(i)}$ a odpovedajúci normalizovaný $i-$ty prvok $\widetilde{v_r}$ označíme*

Iterácie použitím $v_{r+1} = A\widetilde{v_r}$

r-tá iterácia	0	1	2	3	4	5	6
$v_r^{(1)}$	1	8	7.375	7.35593	7.34334	7.35018	7.34608
$v_r^{(2)}$	1	8	7.375	7.33899	7.33642	7.33732	7.33674
$v_r^{(3)}$	1	7	6.375	6.35593	6.34569	6.35112	6.34783
$v_r^{(4)}$	1	6	5.5	5.47458	5.47006	5.47224	5.47091
α_r	1	8	7.375	7.35593	7.34334	7.35018	7.34608
$\widetilde{v_r^{(1)}}$	1	1	1	1	1	1	1
$\widetilde{v_r^{(2)}}$	1	1	1	0.99770	0.99906	0.99825	0.99873
$\widetilde{v_r^{(3)}}$	1	0.87500	0.86441	0.86406	0.86414	0.86408	0.86411
$v_r^{(4)}$	1	0.75000	0.74576	0.74424	0.74490	0.74450	0.74474

$\widetilde{v_r^{(i)}}$.

Iterácie použitím $v_{r+1} = A\widetilde{v_r}$

r-tá iterácia	7	8	9	10
$v_r^{(1)}$	7.34881	7.34748	7.34770	7.34756
$v_r^{(2)}$	7.33797	7.33695	7.33696	7.33695
$v_r^{(3)}$	6.35008	6.34896	6.34913	6.34902
$v_r^{(4)}$	5.47229	5.47137	5.47143	5.47135
α_r	7.34881	7.34748	7.34770	7.34756
$\widetilde{v_r^{(1)}}$	1	1	1	1
$\widetilde{v_r^{(2)}}$	0.99852	0.99857	0.99854	0.99856
$\widetilde{v_r^{(3)}}$	0.86410	0.86410	0.86410	0.86410
$v_r^{(4)}$	0.74465	0.74466	0.74465	0.74465

Toto ukazuje, že po 10 iteráciách aproximácia λ_1 provided by α_1 je $\lambda_1 \approx 7.34756$, a odpovedajúci normalizovaný vlastný vektor je $\widetilde{x_1} \approx [1, 0.99856, 0.86410, 0.74465]^T$. Výpočet s použitím softwarového balíka ukazuje, že aproximácia na 5 desatinných miest dáva výsledok $\lambda_1 = 7.34760$ a $\widetilde{v_1} = [1, 0.99855, 0.86410, 0.74465]^T$.

Rôzna normalizácia, ktorá sa často používa dovoľuje delenie vektora Euklidovou normou vektora $\|u\| = (u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}$, kde $u_1, u_2, ..., u_n$ sú prvky vektora u. Euklidovská norma je vhodná ak pracujeme s vlastnými hodnotami a vlastnými vektormi symetrických matíc, pretože podiel odpovedajúceho terms v postupných iteráciách prepdokladá aproximácie vlastnej hodnoty vyššieho rádu. Metóda mocnín sa dá použiť na hľadanie vlastnej hodnoty λ_n $n \times n$ matice A s najmenšou veľkosťou spolu s vlastným vektorom. .Myšlienka je jednoduchá a je založená na fakte, že ak A je nesingulárna $n \times n$ matica s reálnymi vlastnými hodnotami $\lambda_1, \lambda_2, ..., \lambda_n$, potom sú tieto riešením $Ax = \lambda x$. Pretože je A nesingulárna, má inverznú maticu A^{-1}, a prenásobením rovnice $Ax = \lambda x$ maticou A^{-1} máme $A^{-1}Ax = \lambda A^{-1}x$, alebo $A^{-1}x = \left(\frac{1}{\lambda}\right)x$, čo znamená, že $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda_n}$ sú vlastné hodnoty A^{-1} a že vlastný vektory odpovedajúce λ_i a $\frac{1}{\lambda_i}$ sú identické. Teda ak vlastné hodnoty usporiadame tak, že $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$, vlastná hodnota A s najmenšou magnitúdou bude dominantnou vlastnou hodnotou A^{-1}. Teda aplikácia mocninovej metódy na A^{-1} vygeneruje dominantnú vlastnú hodnotu $\mu_1 = \frac{1}{\lambda_n}$, t.j. $\lambda_n = \frac{1}{\mu_1}$. Ak použijeme túto metódu, inverzná matica A^{-1} sa

nepočíta, a namiesto rovnice

$$Av_{r+1} = v_r \qquad ((45))$$

iterujeme LU dekompozíciu pri riešení v_{r+1} v terms of v_r. Dekompozíciu stačí urobiť iba raz, pretože potom v každom kroku iterácie prvky v_{r+1} môžeme nájsť spätnou substitúciou prvkov v_r. To je situácia, prečo je LU dekompozícia nutná, pretože pravé strany nie sú dané dopredu, preto je nutné riešiť postupnosť problémov s tou istou maticou. Bez LU dekompozície tento proces nie je skutočne praktický. Ako v predchádzajúcej iteračnej procedúre je opäť nutné normalizovať v_r delením každého z jej prvkov prvkom α_r, s najväčšou magnitúdou alebo použiť nejakú inú formu normalizácie, zachovávajúc výpočty vnútri exponent range of the machine. Je to preto, oproti predošlému prípadu kde nenormalizovaný prvok v_r rástol v magnitúde ako rástlo r, v tomto prípade budú klesať, spôsobiac, že presnosť sa stratí ak sa neurobí normalizácia. Táto metóda sa nazýva inverzná mocninová metóda, pretrože je ekvivalentná s iterovaním inverznej matice A^{-1}. Ak označíme normalizovaný stĺpcový vektor v_r ako $\widetilde{v_r}$, môžeme použiť iteračnú schému analogickú s (44)

$$Av_{r+1} = \widetilde{v_r} \quad \text{pre} \quad r = 0, 1, ... \qquad ((46))$$

Example 26 *Použitím inverznej mocninovej metódy nájdite vlastnú hodnotu A s najmenšou magnitúdou, ak* $A = \begin{pmatrix} 4 & 2 & 4 \\ 3 & 9 & 2 \\ 5 & 6 & 9 \end{pmatrix}$.

Solution 27 *Požadovanú vlastnú hodnotu dostaneme iteračným procesom $Av_{r+1} = \widetilde{v_r}$ s danou maticou A, teda sústava ktorú uvažujeme bude*

$$\begin{pmatrix} 4 & 2 & 4 \\ 3 & 9 & 2 \\ 5 & 6 & 9 \end{pmatrix} \begin{pmatrix} v_1^{(r+1)} \\ v_2^{(r+1)} \\ v_3^{(r+1)} \end{pmatrix} = \begin{pmatrix} \widetilde{v_1^{(r)}} \\ \widetilde{v_2^{(r)}} \\ \widetilde{v_3^{(r)}} \end{pmatrix} \qquad s \qquad \begin{pmatrix} v_1^{(0)} \\ v_2^{(0)} \\ v_3^{(0)} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Použitím LU dekompozície dostaneme sústavu:

$4v_1^{(r+1)} + 2v_2^{(r+1)} + 4v_3^{(r+1)} = \widetilde{v_1^{(r)}}$

$\frac{15}{2}v_2^{(r+1)} - v_3^{(r+1)} = \widetilde{v_2^{(r)}}$

$\frac{67}{15}v_3^{(r+1)} = \widetilde{v_3^{(r)}}$

a $v^{(r+1)}$ teraz plynie z $\widetilde{v^{(r)}}$ spätnou substitúciou. S rastúcim r zlomok (podiel) odpovedajúcich komponent $\widetilde{v^{(r+1)}}$ a $\widetilde{v^{(r)}}$ koneguje k vlastnej hodnote μ_1 matice A^{-1} s najväčšou magnitúdou, t.j. vlastnej hodnote A s najmenšou magnitúdou $\lambda_3 = \frac{1}{\mu_1}$. Výsledky po 8 iteráciách sú v tabuľke v predošlom príklade.

Iterácie použitím $Av_{r+1} = \widetilde{v_r}$

Iterácia	0	1	2	3	4
$v_r^{(1)}$	1	0.32090	0.57914	0.61488	0.61215
$v_r^{(2)}$	1	0.02239	−0.12617	−0.16659	−0.17289
$v_r^{(3)}$	1	−0.08209	−0.26606	−0.28158	−0.27571
α_r	1	0.32090	0.57914	0.61488	0.61215
$\widetilde{v_r^{(1)}}$	1	1	1	1	1
$\widetilde{v_r^{(2)}}$	1	0.06977	−0.21786	−0.27093	−0.28243
$\widetilde{v_r^{(3)}}$	1	−0.25582	−0.45941	−0.45794	−0.45040

Iterácie použitím $Av_{r+1} = \widetilde{v_r}$

Iterácia	5	6	7	8
$v_r^{(1)}$	0.60984	0.60898	0.60871	0.60862
$v_r^{(2)}$	−0.17403	−0.17429	−0.17436	−0.17438
$v_r^{(3)}$	−0.27282	−0.27183	−0.27152	−0.27143
α_r	0.60984	0.60898	0.60871	0.60862
$\widetilde{v_r^{(1)}}$	1	1	1	1
$\widetilde{v_r^{(2)}}$	−0.28637	−0.28620	−0.28644	−0.28652
$\widetilde{v_r^{(3)}}$	−0.44736	−0.44637	−0.44606	−0.44598

Odkiaľ vidíme, že aproximatívna hodnota najväčšej vlastnej hodnoty A^{-1} *daná* α_8 *je* $\mu_1 \approx 0.60862$, *teda aproximatívna hodnota najmenšej vlastnej hodnoty* A *je* $\lambda_3 = \frac{1}{\mu_1} = 1.64306$, *a odpovedajúci normalizovaný vlastný vektor* x_3 *daný* v_8 *je* $x_3 \approx [1, -0.28652, -0.44598]^T$. *Výsledky presné na 5 desatinných miest vypočítané softwérovým balíkom sú* $\lambda_3 = 1.64315$ *a* $x_3 = [1, -0.28656, -0.44592]^T$.

Ako možné rozšírenie, nech k je daná konštanta a uvažujme maticu $B = A - kI$. Potom v reči matice B, vlastné hodnoty rovnice

$Ax_i = \lambda_i x_i$ budú

$$Bx_i = (\lambda_i - k)x_i, \qquad ((47))$$

ukazujúc, že vlastné vektory A a B sú identické, ale vlastné hodnoty $\lambda_i - k$ matice B sú vlastné hodnoty matice A redukované o k. To znamená, že vlastné hodnoty $(A - kI)^{-1}$ pre $k = \lambda_i$, s $i = 1, 2, ..., n$, sú $\frac{1}{(\lambda_1 - k)}, \frac{1}{(\lambda_2 - k)}, ..., \frac{1}{(\lambda_n - k)}$. Aplikácia inverznej mocninovej metódy na $(A - kI)^{-1}$ potom určí vlastnú hodnotu A najbližšiu ku konštante k. Túto úvahu môžeme použiť ako základ pre výpočet vlastného vektora a vlastnej hodnoty. Použitím tohto prístupu začiatočná aplikácia inverznej mocninovej metódy je vlastne určovanie vlastnej hodnoty A najbližšej k 0.

Cvičenia.

V cvičeniach 1. - 4. použite mocninovú metódu na hľadanie approximatívnej hodnoty dominantnej vlastnej hodnoty a odpovedajúceho normalizovaného vlastného vektora danej matice štartujúc s $x_0 = [1, 1, 1]^T$ a urobiac 10 iterácií.

1. $A = \begin{pmatrix} 18 & 3 & -1 \\ 3 & 12 & 2 \\ -1 & 2 & 4 \end{pmatrix}$.

2. $A = \begin{pmatrix} 20 & -2 & 1 \\ -2 & 3 & 4 \\ 1 & 4 & 0 \end{pmatrix}$.

3. $A = \begin{pmatrix} 2 & -3 & 2 \\ -3 & 12 & 1 \\ 2 & 1 & 28 \end{pmatrix}$.

4. $A = \begin{pmatrix} -31 & -1 & 2 \\ -1 & -10 & 4 \\ 2 & 4 & -2 \end{pmatrix}$.

V cvičeniach 5. a 6. použite mocninovú metódu na hľadanie approximatívnej hodnoty dominantnej vlastnej hodnoty λ_1, a odpovedajúceho normalizovaného vlastného vektora danej matice štartujúc s $x_0 = [1, -1, 1]^T$ a urobiac 10 iterácií.

5. $A = \begin{pmatrix} 26 & 3 & 1 \\ 3 & 20 & 2 \\ 1 & 2 & 1 \end{pmatrix}$.

6. $A = \begin{pmatrix} 19 & 2 & 2 \\ 2 & 14 & 1 \\ 2 & 1 & 2 \end{pmatrix}$.

V cvičeniach 7. a 10. použite inverznú mocninovú metódu na hľadanie approximatívnej hodnoty vlastnej hodnoty s najmenšou magnitúdou a odpovedajúceho normalizovaného vlastného vektora danej matice štartujúc s $x_0 = [1, 1, 1]^T$ a urobiac 6 iterácií.

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

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

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

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

Numerické riešenie diferenciálnych rovníc.

Väčšina diferenciálnych rovníc nemá známe analytické riešenie, a ak aj nejaké nájdeme, je ťažké ho použiť. Výsledkom je, že ak potrebujeme riešenie a analytické riešenie alebo nie je známe, alebo nie je vhodné na použitie je nutné použiť metódy, ktoré počítajú priamo numerické riešenia. Ale na rozdiel od všeobecného analytického riešenia ZÚ, ktoré možno adaptovať na ľubovoľné začiatočné podmienky, numerické riešenie je riešenie špeciálnej ZÚ, teda výpočet je potrebné zopakovať pri zmene ZP. Sú k dispozícii mnohé rôzne techniky na výpočet numerického riešenia ZÚ, z nich mnohé sú k dispozícii implementované v rôznych numerických balíkoch. V tejto časti sa sústredíme na známe Runge–Kuttove metódy. Metódy typu prediktor–korektor najprv používajú explicitné formuly a dopredu vypočítané riešenia na predikciu nového riešenia. Táto predikcia sa potom zjemňuje použitím implicitnej koretorovej formuly. Metódy Runge–Kutta sú jednokrokové metódy, v ktorých je riešenie diferenciálnej rovnice v nasledujúcom kroku určené výlučne riešením z predchádzajúceho kroku. Na ilustráciu ako numerické riešenie možno získať metódami typu Runge–Kutta a ukázať zmeny stupňa presnosti (aký možno dosiahnuť rôznymi prístupmi) popíšeme niekoľko jednoduchších metód tohto typu.

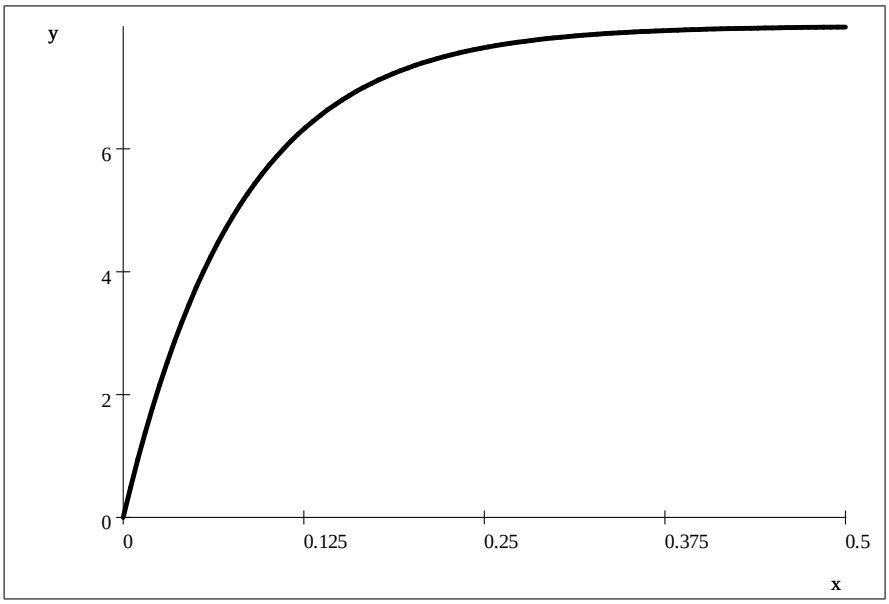
Eulerov algoritmus.

Aproximatívne riešenie ZÚ $\frac{dU}{dt} = f(t, U)$ so ZP $U(0) = 0$ Eulerovou metódou s krokom h is obtained from the algorithm

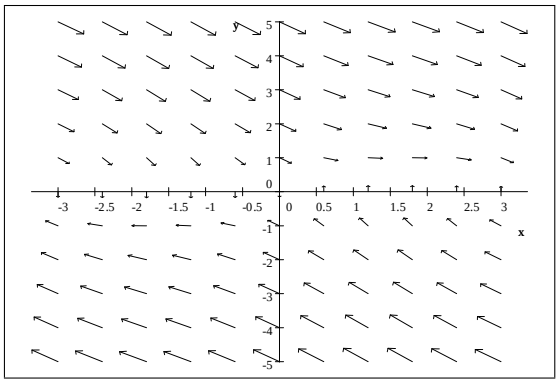
$$U_n = U_{n-1} + hf(t_{n-1}, U_{n-1}), n = 1, 2, ..., \text{ kde } t_n = t_0 + nh.$$

Toto je najjednoduchší príklad jednokrokovej metódy. Obvyklá modifikácia je variácia kroku z jedného bodu do druhého, redukcia, kroku ak sa riešenie rýchlo mení, predlžovanie kroku, ak sa riešenie mení pomaly. Ale nie je možné urobiť také zmeny, systematickým spôsobom, pokiaľ nemáme odhad chyby. Toto obvykle robíme porovnaním výsledku v každom kroku s výsledkom získaným formulou vyššieho rádu.

Example 28 *Použitím Eulerovho algoritmu s krokom* $h = 0.05$ *nájdite približné riešenie ZÚ s diferenciálnou rovnicou prvého rádu* $\frac{dU}{dt} = -12.5U + 100$ *s* $U(0) = 0$ *v intervale* $0 \leq t \leq 0.5$, *a porovnajte ho s presným riešením* $U(t) = 8(1 - e^{-12.50t})$



Solution 29 *Toto je ZÚ, s vektorovým poľom na obrázku:*
$f(t,U) = -12.50U + 100$



Máme $h = 0.05$, $n = 6$, *a* $f(t,U) = -12.50U + 100$ *v Eulerovom algoritme vedie ku nasledujúcim výsledkom. Stĺpec* y_{exact} *obsahuje analytické riešenie.*

n	t_n	U_n	$0.05f(t_n,U_n)$	$U_{n+1} = U_n + 0.05f(t_n,U_n)$
0	0	0	$0.05 \cdot 100 = 5$	5
1	0.05	5	$0.05 \cdot 37.5 = 1.875$	6.875
2	0.1	6.875	$0.05 \cdot 14.063 = 0.703\,15$	$7.578\,2$
3	0.15	$7.578\,2$	$0.05 \cdot 5.272\,5 = 0.263\,63$	$7.841\,8$
4	0.2	$7.841\,8$	$0.05 \cdot 1.977\,5 = 9.887\,5 \times 10^{-2}$	$7.940\,7$
5	0.25	$7.940\,7$	$0.05 \cdot 0.741\,25 = 3.706\,3 \times 10^{-2}$	$7.977\,8$
6	0.3	$7.977\,8$	$0.05 \cdot 0.277\,5 = 1.387\,5 \times 10^{-2}$	$7.991\,7$

$U_{presné}$	$\lvert U_{presné} - U_n \rvert$
0	5
$8\,(1 - e^{-12.5 \cdot 0.05}) = 3.717\,9$	$\lvert 6.875 - 3.7179 \rvert = 3.157\,1$
$8\,(1 - e^{-12.5 \cdot 0.1}) = 5.708\,0$	$\lvert 7.578\,2 - 5.708\,0 \rvert = 1.870\,2$
$8\,(1 - e^{-12.5 \cdot 0.15}) = 6.773\,2$	$\lvert 7.841\,8 - 6.773\,2 \rvert = 1.068\,6$
$8\,(1 - e^{-12.5 \cdot 0.2}) = 7.343\,3$	$\lvert 7.940\,7 - 7.343\,3 \rvert = 0.597\,4$
$8\,(1 - e^{-12.5 \cdot 0.25}) = 7.648\,5$	$\lvert 7.977\,8 - 7.648\,5 \rvert = 0.329\,3$
$8\,(1 - e^{-12.5 \cdot 0.3}) = 7.811\,9$	$\lvert 7.991\,7 - 7.811\,9 \rvert = 0.179\,8$

$u_{2n} = u_{2(n-1)} + hf(t_{n-1}, u_{2(n-1)}), n = 1, 2, ...,$ *kde* $t_n = 0 + 10h$

Solution 30 *Chyba medzi* y_{n+1} *a presným riešením* y_{exact} *môže byť redukovaná, ale nie eliminovaná výberom výberom menšieho kroku, aj keď pre podstatne väčšiu presnosť je nutné použiť inú metódu.*

Eulerova metóda.

Základom tejto metódy je smerové pole odpovedajúce diferenciálnej rovnici prvého rádu

$$\frac{dy}{dx} = f(x, y). \tag{(48)}$$

Pripomeňme definíciu smerového poľa súvisiaceho s (48). V každom bode (x_0, y_0) v (x, y)-rovine v ktorom je $f(x, y)$ definované (48) ukazuje, že sklon (gradient) krivky riešenia v bode je $f(x_0, y_0)$. Ak nakreslíme krátku úsečku v bode (x_0, y_0), ktorá zviera uhol Θ s kladným smerom osi o_x, kde $\operatorname{tg}\Theta = f(x_0, y_0)$, úsečka bude dotyčnicou ku krivke riešenia v (x_0, y_0). Táto úsečka definuje smer zmeny riešenia v bode (x_0, y_0) ak dodáme smerovú šípku, máme smer v ktorom sa y mení v tomto bode pri rastúcom x.Opakovanie tejto konštrukcie v sieti bodov v rovine (x, y), kde je definovaná diferenciálna rovnica (48). Je krátky krok od pojmu smerové pole diferenciálnej rovnice (48) k Eulerovmu algoritmu pre riešenie ZÚ. Aproximatívne numerické riešenie Eulerovou metódou pre ZÚ $\frac{dy}{dx} = f(x, y)$, so začiatočnou podmienkou

$$y(x_0) = y_0, \tag{(49)}$$

dostaneme nasledujúcim spôsobom. Vyberieme dĺžku kroku h premennej x a úsečka cez bod (x_0, y_0) je rozšírená z x_0 do $x_0 + h$, a $y-$ová súradnica $y_0 + y$ koncového bodu úsečky je vybraná ako aproximácia y v $x_0 + h$. Nárast x o h z x_0 spôsobí, že bod na dotyčnicovej aproximačnej úsečke krivky riešenia prechádzajúcej bodom (x_0, y_0) narastie z y_0 na $y_0 + y$, kde $y = h \operatorname{tg}\Theta$, ale $\operatorname{tg}\Theta = f(x_0, y_0)$, teda $y = hf(x_0, y_0)$. Potom máme, že ak P je bod $(x_0 + h, y_1)$ na tangent line approximation,

$$y_1 = y_0 + hf(x_0, y_0). \tag{(50)}$$

Opakovanie tohto procesu produkuje postupnosť bodov $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n), \ldots,$ kde $x_n = x_0 + nh$ a $n = 0, 1, 2, ...$ Ak tieto body spojíme úsečkami, vygenerujeme

polygonálnu line aproximáciu riešenia ZÚ (49), ktorú nazývame Eulerova polygonálna aproximácia riešenia. Algoritmus generovania aproximatívneho riešenia ľahko vidíme:

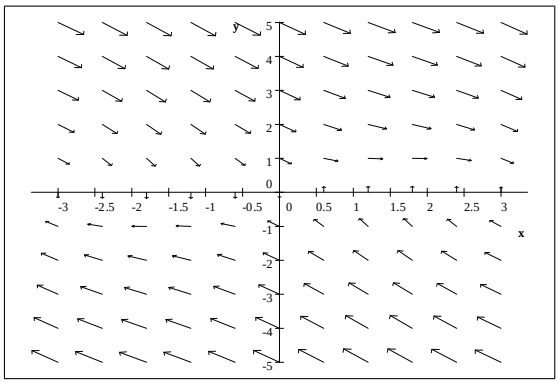
Eulerov algoritmus.
Aproximatívne riešenie ZÚ $\frac{dy}{dx} = f(x, y)$ so ZP $y(x_0) = y_0$ Eulerovou metódou s krokom h dostaneme algoritmus

$$y_n = y_{n-1} + hf(x_{n-1}, y_{n-1}), n = 1, 2, ..., \text{ kde } x_n = x_0 + nh.$$

Toto je najjednoduchší príklad jednokrokovej metódy. Obvyklá modifikácia je variácia kroku z jedného bodu do druhého, redukcia, kroku ak sa riešenie rýchlo mení, predlžovanie kroku, ak sa riešenie mení pomaly. Ale nie je možné urobiť také zmeny, systematickým spôsobom, pokiaľ nemáme odhad chyby. Toto obvykle robíme porovnaním výsledku v každom kroku s výsledkom získaným formulou vyššieho rádu.

Example 31 *Použitím Eulerovho algoritmu s krokom $h = 0.2$ nájdite približné riešenie ZÚ s diferenciálnou rovnicou prvého rádu $\frac{dy}{dx} = sinx - y$ s $y(0) = 1$ v intervale $0 \leq x \leq 2$, a porovnajte ho s presným riešením $y = \frac{1}{2}(\sin x - \cos x) + \frac{3}{2}e^{-x}$.*

Solution 32 *Toto je ZÚ, s vektorovým poľom na obrázku:*
$\sin x - y$



Máme $h = 0,2$, $n = 10$, a $f(x, y) = sinx - y$ v Eulerovom algoritme to vedie k nasledujúcemu výsledku. Stĺpec y_{exact} obsahuje analytické riešenie.

n	x_n	y_n	$0.2f(x_n, y_n)$	$y_{n+1} = y_n + 0.2f(x_n, y_n)$	y_{exact}
0	0	1	-0.2	0.8	1
1	0.2	0.8	-0.1203	0.6797	0.8374
2	0.4	0.6797	-0.0581	0.6217	0.7397
3	0.6	0.6217	-0.0114	0.6103	0.6929
4	0.8	0.6103	0.0214	0.6317	0.6843
5	1	0.6317	0.0420	0.6736	0.7024
6	1.2	0.6736	0.0517	0.7253	0.7366
7	1.4	0.7253	0.0520	0.7773	0.7776
8	1.6	0.7773	0.0444	0.8218	0.8172
9	1.8	0.8218	0.0304	0.8522	0.8485
10	2	0.8522	0.0114	0.8636	0.8657

Chyba medzi y_{n+1} a presným riešením y_{exact} môže byť redukovaná, ale nie eliminovaná výberom menšieho kroku, aj keď pre podstatne väčšiu presnosť je nutné použiť inú metódu.

Modifikovaná Eulerova Metóda.
Zdrojom chýb v EM je jej zlyhanie pri krivosti krivky riešenia v bode (x_i, y_i) keď sa používa dotyčnicová aproximácia pri odhade y_{i+1}. Vylepšenie môžeme dostať použitím dvojstupňového procesu prípravy modifikovaného gradientu $\widetilde{f}(x_i, y_i)$ ktorý môže byť použitý v Eulerovej metóde namiesto $f(x_i, y_i)$. Prvý krok pri hľadaní modifikovaného gradientu zahŕňa výpočet gradientu $f(x_i, y_i)$ a potom ho použiť v EM na výpočet gradientu $f(x_{i+1}, y_{i+1})$ v bode (x_{i+1}, y_{i+1}). Druhý krok zahŕňa priemerovanie týchto dvoch gradientov a výpočet nového gradientu

$$\widetilde{f}(x_i, y_i) = \frac{1}{2}\{f(x_i, y_i) + f(x_{i+1}, y_{i+1})\}, \qquad ((51))$$

a použitie $\widetilde{f}(x_i, y_i)$ namiesto $f(x_i, y_i)$ v EM v bode (x_i, y_i) pri nájdení vylepšeného odhadu $\widetilde{y}_{i+1}$ v bode (x_{i+1}, y_{i+1}) . Tento spôsob výpočtu sa nazýva Heunova metóda, a predstavuje prispevok ku krivosti krivky riešenia v (x_i, y_i) . Nasleduje algoritmus modifikovanej EM:

Aproximatívne numerické riešenie ZÚ $\frac{dy}{dx} = f(x, y)$ so ZP $y(x_0) = y_0$ generované modifikovanou EM s krokom h dostaneme z algoritmu hľadajúceho aproximatívne riešenie modifikovanej EM

$$y_{n+1} = y_n + \frac{1}{2}h[f(x_n, y_n) + f(x_n + h, y_n + hf(x_n, y_n))], \quad \text{pre } n = 1, 2, ..., \text{ kde } x_n = x_0 + nh.$$

Example 33 *Opakujte predošlý príklad použitím modifikovanej EM s $n = 10$ a $h = 0.2$, a porovnajte výsledky získané tak EM ako aj presným riešením.*

Solution 34 *Výsledky výpočtov spolu s porovnaním sú v tabuľke. Detaily nie sú uvedené.*

n	0	1	2	3	4	5
x_n	0	0.2	0.4	0.6	0.8	1
$y_n^{(e)}$	1	0.8	0.6797	0.6217	0.6103	0.6317
$y_n^{(mod)}$	1	0.8399	0.7435	0.6973	0.6887	0.7063
y_{exact}	1	0.8374	0.7397	0.6929	0.6843	0.7024
n	6	7	8	9	10	
x_n	1.2	1.4	1.6	1.8	2	
$y_n^{(e)}$	0.6736	0.7253	0.7773	0.8212	0.8522	
$y_n^{(mod)}$	0.7397	0.7796	0.8181	0.8482	0.8643	
y_{exact}	0.7366	0.7776	0.8172	0.8485	0.8657	

Porovnanie výsledkov v posledných troch riadkoch ukazuje zlepšenie v presnosti pre modifikovanú EM.

EM je skutočne Taylorov rad riešenia $y(x)$, v ktorom $y\,(x_n + h)$ je predikovaný z $y(x_n)$ použitím dvoch členov Taylorovho radu funkcie $y(x)$ v bode x_n.

Často používanou numerickou metódou integrácie ZÚ pre rovnice prvého rádu je Runge–Kutta metóda štvrtého rádu. Existuje niekoľko typov formúl štvorkrokovej Runge Kutta metódy štvrtého rádu, v ktorých chyba po použití kroku h je rádu h^5, ale my popíšeme najznámejšiu. Najskôr ukážeme všeobecný prístup k odvodeniu R-KM použitím modifikovanej EM. Základom každej R-KM sú jednokrokové metódy, ktoré môžme uvažovať v tvare

$$y_{i+1} = y_i + hF(x_i, y_i, h), \qquad ((52))$$

kde $F(x_i, y_i, h)$ reprezentuje istý tvar priemernej hodnoty $f(x, y)$ na intervale $x_i \leq x \leq x_{i+1}$. Všetky tieto metódy možno získať úpravou partikulárnej formy F, ktorá obsahuje nejaké neurčené konštanty a potom výpočtom rovníc určujúcich konštanty vyžadujeme, aby F súhlasila s Taylorovým radom f do určitého stupňa h. V prípade ak F obsahuje termy rádu h, tak chyba v každom kroku bude rádu h^2, použitím reťazového pravidla a faktu, že $f(x, y) = f(x, y(x))$, funkcia F v (52) je aproximovaná useknutým Taylorovým radom R–K typ derivácie modifikovanej EM

$$F(x, y, h) = f(x, y) + \frac{1}{2}h\left\{\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}\right\},$$

ale $\frac{dy}{dx} = f(x, y)$, teda

$$F(x, y, h) = f(x, y) + \frac{1}{2}h\left\{f_x(x, y) + f_y(x, y)f(x, y)\right\}. \qquad ((53))$$

Teraz nájdeme reprezentáciu funkcie F v tvare

$$F(x, y, h) = w_1 f(x, y) + w_2 f(x + w_3 h, y + w_4 h f(x, y)), (54) \qquad ((54))$$

kde dosiaľ konštanty w_1 až w_4 sú neznáme. Rozkladom $f(x + w_3 h, y + w_4 h f(x, y))$ v bode (x, y) ako Taylorovho radu dvoch premenných so zvyškom po prvých deriváciách dáva

$$\begin{aligned} &f(x + w_3 h, y + w_4 h(f(x, y)) = \\ &= f(x, y) + w_3 h f_x(x, y) + w_4 h f_y(x, y)f(x, y) + R(h), \end{aligned} \qquad ((55))$$

kde chyba $R(h)$ je rádu h^2. Dosadením (55) do (54) a kombináciou máme

$$F(x, y, h) = (w_1 + w_2)f(x, y) + h(w_2 w_3 f_x(x, y) + w_2 w_4 f_y(x, y)f(x, y)). \qquad ((56))$$

Ak majú byť (54) a (56) rovnaké až do výrazov rádu h, porovnaním výrazov odpovedajúcich jednotlivým mocninám h dostaneme $w_1 + w_2 = 1, w_2 w_3 = \frac{1}{2}$, a $w_2 w_4 = \frac{1}{2}$.

Tieto tri rovnice spájajú konštanty w_1 až w_4, teda ak jedna z konštánt napríklad w_2, je ľubovoľne vybratá, ostatné sa určia pomocou nej. Z (54) potom máme

$$F(x, y, h) = (1 - w_2)f(x, y) + w_2 f\left(x + \frac{1}{2}\frac{h}{w_2}, y + \frac{1}{2}\frac{hf(x, y)}{w_2}\right). \qquad ((57))$$

Ak napríklad $w_2 = \frac{1}{2}$ v (57), a použitím (52), máme modifikovanú EM

$$y_{i+1} = y_i + \frac{1}{2}h\{f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))\}. \qquad ((58))$$

CARL DAVID TOLME RUNGE (1856–1927) Nemecký matematik, profesor aplikovanej matematiky v Göttingene. Zaoberal sa numerickým riešením diferenciálnych rovníc, jeho prístup aplikoval Wilhelm Kutta (1867–1944), nemecký aerodynamik, ktorý použil Rungeho prácu pri štúdiu mechaniky tekutín.

R–KM štvrtého rádu pre diferenciálnu rovnicu prvého rádu.
Aproximácia numerického riešenia ZÚ $\frac{dy}{dx} = f(x, y)$ so ZP $y(x_0) = y_0$ s krokom dĺžky h dostaneme z nasledujúceho R-K algoritmu štvrtého rádu, s $x_n = x_0 + nh$ a $y_n = y(x_n)$.

KROK 1 Výpočet

$$k_{1n} = hf(x_n, y_n)$$
$$k_{2n} = hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_{1n}\right)$$
$$k_{3n} = hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_{2n}\right)$$
$$k_{4n} = hf(x_n + h, y_n + k_{3n}).$$

KROK 2 Výpočet

$$d_n = \frac{1}{6}(k_{1n} + 2k_{2n} + 2k_{3n} + k_{4n}).$$

KROK 3 Numerická aproximácia y_{n+1} riešenia $y = y(x_{n+1})$ daná

$$y_{n+1} = y_n + d_n, \quad n = 1, 2, ...$$

Example 35 *Použitím Runge-Kutta algoritmu štvrtého rádu s krokom* $h = 0.2$ *riešte ZÚ* $\frac{dy}{dx} + 2y = \sin 3x$ *s* $y(0) = 1$ *v intervale* $0 \le x \le 2.4$. *Porovnajte výsledok s výsledkom získaným modifikovanou EM a s analytickým riešením*
$y = \frac{1}{13}[9\cos x - 2\sin x + 4\sin 2x\cos x - 12\cos^3 x + 16e^{-2x}]$.

Solution 36 *V nasledujúcich výpočtoch* $f(x, y) = \sin 3x - 2y$, *krok dĺžky* $h = 0.2$, *teda ak riešenie hľadáme v intervale* $0 \le x \le 2.4$, *tak* $n = 0, 1, ..., 12$. *Detaily vnútorných výpočtov pre* $x = 0, 0.2, 0.4$ *a* 0.6 *sú v prvej z nasledujúcich tabuliek. Under the heading yrk, the second table lists all of the results obtained by the Runge–Kutta algorithm up to x =2.4, and for purposes of comparison the columns with headings ymod and yexact show the results obtained by using the modified Euler method and the analytical solution, respectively.*

Detailné výpočty pre $x = 0, 0.2,$ *a* 0.4

n	x_n	y_n	$f(x_n, y_n)$	k_{1n}	k_{2n}
0	0	1	-2	-0.4	-0.2609
1	0.2	0.72153	-0.87842	-0.17568	-0.09681
2	0.4	0.61292	-0.29380	-0.05876	-0.03392
3	0.6	0.57305	$-$	$-$	$-$

n	x_n	y_n	k_{3n}	k_{4n}	y_{n+1}
0	0	1	-0.28872	-0.17158	0.72153
1	0.2	0.72153	-0.11258	-0.05717	0.61292
2	0.4	0.61292	-0.03889	-0.03484	0.57305
3	0.6	0.57305	$-$	$-$	$-$

Porovnanie výsledkov v intervale $0 \leq x \leq 2.4$

n	x_n	y_{rk}	y_{mod}	y_{exact}	n	x_n	y_{rk}	y_{mod}	y_{exact}
0	0	1.0	1.0	1.0	7	1.4	0.05390	0.05090	0.05389
1	0.2	0.72153	0.73646	0.72142	8	1.6	−0.12324	−0.11730	−0.12328
2	0.4	0.61292	0.62788	0.61279	9	1.8	−0.23165	−0.21681	−0.23173
3	0.6	0.57305	0.58026	0 .57295	10	2.0	−0.24192	−0.22174	−0.24202
4	0.8	0.52262	0.52056	0.52257	11	2.2	−0.15615	−0.13639	−0.15624
5	1.0	0.41675	0.40862	0.41674	12	2.4	−0.00809	−0.00531	−0.00816
6	1.2	0.25051	0.24208	0.25052	−	−	−	−	−

R–KM štvrtého rádu je ľahko adaptovateľná na riešenie sústavy dvoch diferenciálnych rovníc prvého rádu, alebo ako špeciálny príklad na riešenie jednej diferenciálnej rovnice druhého rádu:

R–KM štvrtého rádu pre sústavu dvoch diferenciálnych rovníc prvého rádu.

Aproximácia numerického riešenia ZÚ pre sústavu $\frac{dy}{dx} = f(x, y, z)$ a $\frac{dz}{dx} = g(x, y, z)$ so ZP $y(x_0) = y_0$ a $z(x_0) = z_0$ s krokom dĺžky h dostaneme z nasledujúceho R-K algoritmu štvrtého rádu s krokom dĺžky h, kde $x_n = x_0 + nh, y_n = y(x_n)$ a $zn = z(x_n)$.

KROK 1 Výpočet v nasledujúcom poradí

$$k_{1n} = hf(x_n, y_n, z_n) \quad K_{1n} = hg(x_n, y_n, z_n)$$

$$k_{2n} = hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_{1n}, z_n + \frac{1}{2}K_{1n}\right) \quad K_{2n} = hg\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_{1n}, z_n + \frac{1}{2}K_{1n}\right)$$

$$k_{3n} = hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_{2n}, z_n + \frac{1}{2}K_{2n}\right) \quad K_{3n} = hg\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_{2n}, z_n + \frac{1}{2}K_{2n}\right)$$

$$k_{4n} = hf\left(x_n + h, y_n + k_{3n}, z_n + K_{3n}\right) \quad K_{4n} = hg\left(x_n + h, y_n + k_{3n}, z_n + K_{3n}\right).$$

KROK 2 Výpočet

$$d_n = \frac{1}{6}(k_{1n} + 2k_{2n} + 2k_{3n} + k_{4n}) \text{ a } D_n = \frac{1}{6}(K_{1n} + 2K_{2n} + 2K_{3n} + K_{4n}).$$

KROK 3 Numerická aproximácia riešení $y = y(x_{n+1})$ a $z = z(x_{n+1})$ je daná $y_{n+1} = y_n + d_n$ a $z_{n+1} = z_n + D_n$, pre $n = 1, 2, ...$

Tento Runge–Kutta algoritmus štvrtého rádu s krokom h možno ľahko modifikovať na hľadania riešenia nasledujúcich ZÚ pre jednu diferenciálnu rovnicu druhého rádu. Štandardná forma pre adaptáciu R–KM na riešenie rovnice druhého rádu

$$\frac{d^2y}{dx^2} = g\left(x, y, \frac{dy}{dx}\right) \quad \text{s } y(x_0) = y_0 \text{ a } z(x_0) = z_0. \tag{(59)}$$

Všetko čo je nutné na redukciu d.r. druhého rádu na sústavu dvoch d.r. prvého rádu je, že položíme

$$\frac{dy}{dx} = z \text{ a } \frac{dz}{dx} = g(x, y, z) \tag{(60)}$$

v prechádzajúcom R-K algoritme, a potom použijeme podmienky

$$y(x_0) = y_0 \text{ a } z(x_0) = y(x_0) = z_0. \tag{(61)}$$

Example 37 *Použitím R-K algoritmu s krokom dĺžky* 0.1 *nájdite numerickú aproximáciu riešenia ZÚ pre Hermiteovu rovnicu* $y'' - 2xy' + 8y = 0$ *s ZP* $y(0) = 12$ *a* $y'(0) = 0$ *v intervale* $0 \leq x \leq 1$. *Porovnajte výsledok výpočtov s analytickým riešením* $y(x) = 16x^4 - 48x^2 + 12$.

Solution 38 *Toto je Hermitova rovnica s* $n = 4$, *s analytickým riešením* $H_4(x) = 16x^4 - 48x^2 + 12$. *Použitím (59) a (60) položme* $z = \frac{dy}{dx}$ *a* $g(x, y, z) = 2xz - 8y$, *a použijeme krok dĺžky* $h = 0.1$. *ZP sú dané v začiatku, teda* $x_0 = 0, y(x_0) = 12$ *a* $z(x_0) = y(x_0) = 0$, *odpovedajúco* $y_0 = 12$ *a* $z_0 = 0$. *Uvádzame detaily vnútorných výpočtov pre* $x = 0$ *a* 0.1; *nasledujúca tabuľka výsledkov pre interval* $0 \leq x \leq 1$, *s R-K riešením rovnice druhého rádu označenom* y_{rk} *a analytickým riešením* y_{exact}.

$x_0 = 0$

$f(x_0, y_0, z_0) = 0$, $g(x_0, y_0, z_0) = -96$, $k_1 = 0$, $K_1 = -9.6$, $k_2 = -0.48$, $K_2 = -9.648$, $k_3 = -0.4824$, $K_3 = -9.45624$, $k_4 = -0.945624$, $K_4 = -9.403205$, $d = -0.478404$, $D = -9.535281$, *teda* $y_1 = 11.521596$ *a* $z_1 = -9.535281$, *kde* $z_1 = y'(x_1)$.

$x_1 = 0.2$

$f(x_1, y_1, z_1) = -9.535281$, $g(x_1, y_1, z_1) = -94.079824$, $k_1 = -0.953528$, $K_1 = -9.407982$, $k_2 = -1.423927$, $K_2 = -9.263044$, $k_3 = -1.416680$, $K_3 = -9.072710$, $k_4 = -1.860799$, $K_4 = -8.828252$, $d = -1.415924$, $D = -9.151290$, *teda* $y_2 = 10.105672$ *a* $z_2 = -18.686571$, *kde* $z_2 = y'(x_2)$.

Porovnanie riešení pre $0 \leq x \leq 1$

n	x_n	y_{rk}	y_{exact}	n	x_n	y_{rk}	y_{exact}
0	0	12	12	6	0.6	−3.205311	−3.2064
1	0.1	11.521596	11.5216	7	0.7	−7.676938	−7.6784
2	0.2	10.105672	10.1056	8	0.8	−12.164555	−12.1664
3	0.3	7.809827	7.8096	9	0.9	−16.380188	−16.3824
4	0.4	4.730055	4.7296	10	1.0	−19.997470	−20.0
5	0.5	1.000747	1.0	–	–	–	–

Keď sa riešenie diferenciálnej rovnice mení rýchlo v nejakých intervaloch a pomaly v iných intervaloch, je nutné meniť dĺžku kroku pri výpočtoch za predpokladu že zachováme presnosť. R-K algoritmus založený na tvare R–K schémy štvrtého rádu sa zvykne implementovať v mnohých numerických softwarových programoch, ktoré sú schopné meniť dĺžku kroku v každom štádiu výpočtu. Nárast v ztložitosti výpočtov sa indikuje faktom, že algoritmus používa adaptivnu dĺžku kroku v šiestich etapách výpočtu namiesto štyroch používaných klasickým R-K algoritmom. ako výpočty postupujú, numerické odhady riešenia po danej dĺžke kroku h sa robia použitím tvaru R-KM štvrtého rádu a účinnej formuly piateho rádu. Rozdiel medzi týmito dvomi odhadmi sa porovná s dopredu zadanou toleranciou a výsledok sa potom použije s redukovaným, alebo s rastúcim krokom pokiaľ rozdiel leží vnútri požadovanej tolerancie. Výsledná dĺžka kroku sa použije s výhodou pri výpočte v nasledujúcom štádiu výpočtu.

Cvičenia.

1. Riešte nasledujúce ZÚ použitím R-K algoritmu štvrtého rádu.

2. $y' = (3x^2 + y^2)^{\frac{1}{2}} - y$ s $y(2) = 0$ a $h = 0.2$ na intervale $2 \leq x \leq 3$,

3. $y' = \frac{xy}{(x^2+y^2)^{\frac{1}{2}}}$ s $y(1) = 1$ a $h = 0.2$ na na intervale $1 \leq x \leq 2$.

4. $y' = (x^2 + y^2)^{\frac{1}{2}} - xy$ s $y(1) = 2$ a $h = 0.2$ na intervale $1 \leq x \leq 2$.

5. $y' = \frac{1}{2}(x^2 + 2y^2) - xy$ s $y(1) = 0$ a $h = 0.1$ na intervale $1 \leq x \leq 1.5$.

6. $y' = \cos(2x + y) - 3y$ s $y(1) = 1$ a $h = 0.2$ na intervale $1 \leq x \leq 2$.

7. $y' = \sin(x + y) - 2y$ s $y(0) = 1$ a $h = 0.2$ na intervale $0 \leq x \leq 1$.

8. $y'' - xyy' + 2y = 0$ s $y(0) = 2,\ y'(0) = 2$ a $h = 0.1$ na intervale $0 \leq x \leq 0.5$.

9. $y'' + (3 + x)y' + y^2 = 0$ s $y(1) = 1,\ y'(1) = 2$ a $h = 0.1$ na intervale $1 \leq x \leq 1.5$.

10. $y'' + (1 + \sin 2x)y' + 3y = 0$ s $y(0) = 1,\ y'(0) = 1$ a $h = 0.1$ na intervale $0 \leq x \leq 0.5$.

11. $y'' + (1 + y^2)^{\frac{1}{2}}y' + y = 0$ s $y(2) = 0,\ y'(2) = 1$ a $h = 0.1$ na intervale $2 \leq x \leq 2.5$.

12. $y'' + 2y' - y^2 = 0$ s $y(0) = 2,\ y'(0) = 1$ a $h = 0.2$ na intervale $0 \leq x \leq 1$.

13. $y'' - xy' - y^2 = 0$ s $y(0) = -1,\ y'(0) = 2$ a $h = 0.2$ na intervale $0 \leq x \leq 1$.

14. $y'' + yy' - 3y = 0$ s $y(1) = 1,\ y'(1) = 1$ a $h = 0.2$ na intervale $1 \leq x \leq 2$.

15. $y'' + x2\sin y' - 2y = 0$ s $y(1) = 0,\ y'(0) = -1$ a $h = 0.2$ na intervale $1 \leq x \leq 2$.

16. $y'' - xy' - y^2 = 2x$ s $y(0) = -2,\ y'(0) = 1$ a $h = 0.2$ na intervale $0 \leq x \leq 1$.

17. $y'' + 2yy' - 3y = 1 - x^2$ s $y(0) = 3,\ y'(0) = 2$ a $h = 0.2$ na intervale $0 \leq x \leq 1$.

18. $\frac{dx}{dt} = tx - (x + y)y$ a $\frac{dy}{dt} = ty - (x + y)x$ s $x(0) = 1,\ y(0) = 0$ a $h = 0.2$ na intervale $0 \leq t \leq 1$.

19. $\frac{dx}{dt} = (1 + t)y^2 - 2x$ a $\frac{dy}{dt} = y^2 + tx$ s $x(0) = -1,\ y(0) = -3$ a $h = 0.2$ na intervale $0 \leq t \leq 1$.

20. $\frac{dx}{dt} = \sin(x + 4y)$ a $\frac{dy}{dt} = 2\cos(x - 3y)$ s $x(0) = 1,\ y(0) = 1$ a $h = 0.2$ na intervale $0 \leq t \leq 1$.

21. $\frac{dx}{dt} = \sin x + 4\cos y$ a $\frac{dy}{dt} = \sin y - 3\sin x$ s $x(0) = 1,\ y(0) = -2$ a $h = 0.2$ na intervale $0 \leq t \leq 1$.

BIBLIOGRAPHY

[1] Galanová, J., Gatial, J., Kaprálik, P.: Lineárna algebra, STU Bratislava, 2002

[2] Marko Ľ.: Matematická analýza online, 2001, http://aladin.elf.stuba.sk/~marko

[3] Stroud,K.A.: Engineering Mathematics, Macmillan Presss LTD, Hong Kong, 1993

[4] Šulka, R., Moravský, L., Satko, L.: Matematická analýza I, Alfa, SNTL, Bratislava 1986

[5] Glyn, J.: Modern engineering mathematics, Addison Wesley, 2008