NUMERICKÉ METODY

Josef Dalík

1 Zdroje chyb

Při řešení daného technického problému numerickými metodami jde zpravidla o zjištění některých kvantitativních charakteristik daného procesu probíhajícího v přírodě nebo v některém technickém systému. Může jít například o tok kapaliny, chemickou reakci nebo o deformaci tělesa.

a) *Matematický model úlohy* je exaktní matematický popis podstatných vztahů, které daný proces určují. Je tvořen například okrajovou diferenciální úlohou, soustavou rovnic nebo integrálem.

b) *Diskretizace modelu úlohy (Numerická úloha)* musí poskytnout přesné vymezení konečného souboru vstupních dat, konečného souboru výstupních dat a algoritmu, který k daným vstupním datům přiřadí jediná výstupní data.

Definice. Nechť $\hat{x}$ je hledaná přesná veličina a x její aproximace. Pak

$$e_x = \hat{x} - x$$

se nazývá *(absolutní) chyba aproximace* a $\varepsilon(x)$ s vlastností

$$|e_x| \leq \varepsilon(x)$$

se nazývá *odhad chyby*. Píšeme $\hat{x} = x \pm \varepsilon(x)$ a tento výraz nazýváme *neúplné číslo*. Zlomek

$$\frac{e_x}{x}$$

se nazývá *relativní chyba aproximace* a $\delta(x)$,

$$\left|\frac{e_x}{x}\right| \leq \delta(x)$$

se nazývá *odhad relativní chyby*.

Definice. Je-li

a) $\hat{x}$ přesné řešení daného technického problému a
x přesné řešení matematického modelu,
pak e_x se nazývá *chyba matematického modelu*.

b) $\hat{x}$ přesné řešení matematického modelu a

x přesné řešení diskretizace,

pak e_x se nazývá *chyba aproximace (chyba numerické metody).*

c) $\hat{x}$ přesné řešení diskretizace a

x skutečně vypočtená aproximace řešení,

pak e_x se nazývá *zaokrouhlovací chyba.*

d) $\hat{x}$ přesné řešení daného technického problému a

x skutečně vypočtená aproximace řešení,

pak je e_x v jistém smyslu součtem chyb všech výše uvedených typů.

2 Startovací metody pro jednu rovnici s jednou reálnou neznámou

Pro daný interval I v $\mathbf{R}$ a pro reálnou funkci f na I najděte řešení rovnice

$$f(x) = 0. \tag{1}$$

Číslo x se nazývá *kořen rovnice* (1).

Příklad 1. Určete přibližně všechny kořeny rovnice

$$f(x) \equiv 10 \sin x - x - 5 = 0$$

z intervalu $(0, \pi)$.

Představu o počtu a přibližný odhad hodnot kořenů poskytne tato *grafická metoda:*

$$f(x) = 0 \Longleftrightarrow 10 \sin x = x + 5$$

a z grafického znázornění lze odhadnout, že rovnice má dva kořeny $x_1 \doteq 0.5$ a $x_2 \doteq 2.4$. Přesnější informaci poskytne tato zkouška:

x	f(x)
0.5	-0.7057
0.7	0.7422
2.4	-0.6454
2.2	0.8850

Fakt, že hodnoty funkce f v bodech 0.5 a 0.7 mají opačná znaménka říká, že kořen x_1 leží v intervalu $(0.5, 0.7)$ a zbývající dva řádky vedou k závěru, že kořen x_2 leží v intervalu $(2.2, 2.4)$. V této úvaze bylo využito této základní vlastnosti funkcí spojitých na uzavřeném intervalu:

Věta 1. Jestliže $f \in C\langle a, b\rangle$ a platí-li $f(a) \cdot f(b) < 0$, pak v otevřeném intervalu (a, b) leží alespoň jeden kořen rovnice $f(x) = 0$.

Na opakovaném použití Věty 1 jsou založeny dvě níže uvedené metody A), B). Jsou použitelné za předpokladu, že je dán interval (a_0, b_0) tak, že

$$f \in C\langle a_0, b_0\rangle \quad \text{a} \quad f(a_0) \cdot f(b_0) < 0.$$

Spočívají v tom, že se konstruuje posloupnost intervalů

$$(a_0, b_0) \supset (a_1, b_1) \supset \ldots \supset (a_n, b_n) \supset \ldots$$

tak, aby $f(a_n) \cdot f(b_n) < 0$ pro $n = 1, 2, \ldots$ Podinterval (a_n, b_n) se v intervalu (a_{n-1}, b_{n-1}) najde takto: Vybere se bod $s_n \in (a_{n-1}, b_{n-1})$. Protože $f(a_{n-1}) \cdot f(b_{n-1}) < 0$, platí právě jedna z podmínek

1. $f(a_{n-1}) \cdot f(s_n) < 0 \ldots$ položíme $a_n = a_{n-1}$, $b_n = s_n$
2. $f(s_n) \cdot f(b_{n-1}) < 0 \ldots$ položíme $a_n = s_n$, $b_n = b_{n-1}$
3. $f(s_n) = 0 \ldots$ výpočet skončí.

Metody A), B) se liší způsobem výpočtu bodu s_n a kritériem pro ukončení.

A)Metoda půlení intevalu

$$s_n = \frac{a_{n-1} + b_{n-1}}{2}$$

a pro předem dané malé kladné číslo ε výpočet skončí, jakmile $|a_n - b_n| < \varepsilon$.

Pro metodu půlení intervalu lze velmi snadno posoudit rychlost konvergence. Na začátku výpočtu splňuje kořen x rovnice podmínku

$$x = s_1 \pm d_0 \qquad d_0 = \frac{b_0 - a_0}{2}$$

a po n krocích je

$$x = s_{n+1} \pm d_n, \quad d_n = \frac{b_n - a_n}{2} = \ldots = \frac{b_0 - a_0}{2^{n+1}}, \quad d_n = \frac{d_0}{2^n}$$

Tedy v každém kroku se odhad chyby zmenší dvakrát.

Příklad 2. Kolik kroků metody půlení intervalu je třeba k tomu, aby se odhad chyby zmenšil desetkrát?

Protože po n krocích se odhad chyby zmenší 2^n krát, hledáme nejmenší hodnotu n tak, aby $10 < 2^n$. Protože $10 \doteq 2^{3,3}$, je pro zmenšení chyby desetkrát třeba provést čtyři kroky metody půlení intervalu.

Příklad 3. Kolik kroků metody A) poskytne kořen x^1 z příkladu 1 s chybou menší než 10^{-3}?

$$x^1 \in (0,5, 0,7) \Longrightarrow x^1 = 0,6 \pm 0,1 \quad \text{a} \quad d_0 = 0,1$$

Hledáme tedy n tak, aby

$$\frac{0,1}{2^n} \leq 10^{-3} \Longleftrightarrow 2^n \geq 100 \Longleftrightarrow n \geq 7$$

Je tedy nutno provést alespoň 7 kroků metody půlení. Doporučený způsob záznamu výpočtu je ilustrován v této tabulce:

n	a_{n-1}	$f(a_{n-1})$	b_{n-1}	$f(b_{n-1})$	s_n	$f(s_n)$
1	0,5	-	0,7	+	0,6	+
2	0,5	-	0,6	+	0,55	-
3	0,55	-	0,6	+	0,575	-
4	0,575	-	0,6	+	0,5875	-
5	0,5875	-	0,6	+	0,59375	+
6	0,5875	-	0,59375	+	0,590625	-
7	0,590625	-	0,59375	+	0,5921875	-
8	0,5921875	-	0,59375	+	0,59296875	-

Tedy $x^1 = 0,59296875 \pm 10^{-3}$, přesněji $x^1 = 0,59296875 \pm 0.00078125$.

B)Metoda regula falsi

Nový bod s_n je průsečík osy x s přímkou, spojující body $[a_{n-1}, f(a_{n-1})]$, $[b_{n-1}, f(b_{n-1})]$, tj.

$$s_n = a_{n-1} - f(a_{n-1})\frac{b_{n-1} - a_{n-1}}{f(b_{n-1}) - f(a_{n-1})}.$$

Kriterium pro ukončení: Zvolí se $\delta > 0$ a výpočet skončí, jakmile $|f(s_n)| \leq \delta$.

Příklad 4. Najděte kořen rovnice

$$10 \sin x - x - 5 = 0$$

na intervalu $(0,5, 0,7)$ s tolerancí $\delta = 0,0002$.

n	a_{n-1}	b_{n-1}	$f(a_{n-1})$	$f(b_{n-1})$	s_n	$f(s_n)$
1	0,5	0,7	-0,705745	0,742177	0,597484	0,028156
2	0,5	0,597484	-0,705745	0,028156	0,5973744	0,000938
3	0,5	0,593744	-0,705745	0,000938	0,593619	0,000031

Tedy $x^1 \doteq 0,59361957$.

3 Princip metody postupných aproximací (Princip iteračních metod)

Příklad 1. Uvažme rovnici

$$e^{-x} - 2x = 0.$$

Její ekvivalentní vyjádření $e^{-x} = 2x$ má tu výhodu, že grafy funkcí e^{-x} a $2x$ jsou dobře známé. Z jejich schematického znázornění lze usoudit, že rovnice má jediné řešení $\hat{x}$ a $\hat{x} \doteq 0,303$.

Řešení dané rovnice iterací:

a) $x = -\ln(2x)$:

$x_0 = 0.303 \qquad x_{n+1} = -\ln(2x_n)$

n	x_n
0	0,303
1	0,501
2	-0,002
3	–

b) $x = \frac{1}{2}e^{-x}$:

$x_0 = 0.303 \qquad x_{n+1} = \dfrac{1}{2}e^{-x_n}$

n	x_n
0	0,303
1	0,369
2	0,346
⋮	⋮
6	0,352
7	0,352

Definice. Nechť $X \neq \emptyset$ a $F : X \longrightarrow X$. Prvek $x \in X$ se nazývá *pevný bod* zobrazení F, když $x = F(x)$.

Metoda postupných aproximací hledá pevný bod zobrazení F tak, že se x_0 zvolí a pro $n = 0, 1, \ldots$ se postupně počítá $x_{n+1} = F(x_n)$. Předpokládejme, že $\lim_{n\to\infty} x_n = x$. Pak, je-li zobrazení F spojité, platí

$$x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} F(x_n) = F(\lim_{n\to\infty} x_n) = F(x).$$

Definice. Nechť $X \neq \emptyset$ a ke každým prvkům $x, y \in X$ je přiřazeno reálné číslo $d(x, y)$ tak, že

D1 $d(x, y) \geq 0, \qquad d(x, y) = 0 \Leftrightarrow x = y$

D2 $d(x, y) = d(y, x)$

D3 $d(x, y) \leq d(x, z) + d(z, y)$

Pak X se nazývá *metrický prostor*, prvky z X se nazývají *body* a funkce d *metrika (vzdálenost)* v X.

Definice. Nechť $(x_n)_{n=1}^{\infty}$ je posloupnost bodů v metrickém prostoru X a $x \in X$. Položíme

$$x = \lim_{n\to\infty} x_n,$$

jestliže $d(x_n, x) \longrightarrow 0$ pro $n \longrightarrow \infty$ [ke každému $\varepsilon > 0$ existuje $n_0 > 0$: $d(x_n, x) < \varepsilon$ pro všechna $n > n_0$].

Každá posloupnost, která má v X limitu, se nazývá *konvergentní*.

Věta 1. Posloupnost bodů metrického prostoru může mít nejvýše jednu limitu.

Definice. Posloupnost $(x_n)_{n=1}^{\infty}$ bodů metrického prostoru X se nazývá *cauchyovská*, jestliže

$$d(x_k, x_l) \longrightarrow 0 \quad \text{pro} \quad k, l \longrightarrow \infty$$

[ke každému $\varepsilon > 0$ existuje $n_0 > 0$ tak, že

$$d(x_n, x_{n+p}) < \varepsilon \quad \text{pro všechna} \quad n > n_0, p > 0$$

Věta 2. Každá konvergentní posloupnost v metrickém prostoru je cauchyovská.

Obecně není pravda, že každá cauchyovská posloupnost je konvergentní.

Definice. Metrický prostor X se nazývá *úplný*, je-li každá cauchyovská posloupnost v X konvergentní.

ÚLOHA. Buďte X metrický prostor a $F : X \longrightarrow X$. Hledáme $x \in X$ splňující

$$x = F(x).$$

Definice. Nechť X je metrický prostor, $F : X \longrightarrow X$ a $0 \leq \alpha < 1$. Zobrazení F se nazývá *kontrakce* v X *s koeficientem* α, jestliže

$$d(F(x), F(y)) \leq \alpha d(x, y) \quad \text{pro všechna} \quad x, y \in X.$$

Věta 3. (Věta o kontrakci, Banachova věta o pevném bodu) Nechť X je úplný metrický prostor, F kontrakce v X s koeficientem α, $x_0 \in X$ je libovolný bod a $(x_n)_{n=1}^{\infty}$ je příslušná posloupnost postupných aproximací. Pak platí

a) V X existuje jediný pevný bod $\hat{x}$ zobrazení F.

b) $\hat{x} = \lim_{n\to\infty} x_n$.

c) $d(x_n, \hat{x}) \leq \alpha^n d(x_0, \hat{x})$ pro $n = 1, 2, \ldots$

d) $d(x_n, x_0) \leq \frac{\alpha^n}{1-\alpha}\, d(x_1, x_0)$ pro $n = 1, 2, \ldots$

Tvrzení věty 3

a) zodpovídá otázku existence a jednoznačnosti řešení ÚLOHY,

b) popisuje postup přibližného řešení ÚLOHY,

c) říká, že x_n je tím blíže k $\hat{x}$, čím

- blíže k $\hat{x}$ je x_0,
- menší je koeficient α,
- větší je index n

d) je prakticky použitelný odhad chyby.

Důkaz věty 3. 1. JEDNOZNAČNOST: Předpokládejme, že $u = F(u)$ a $v = F(v)$. Pak užitím D1 dostaneme

$$d(u,v) = d(F(u),F(v)) \leq \alpha d(u,v) \implies (1-\alpha)d(u,v) \leq 0 \implies$$
$$d(u,v) \leq 0 \implies d(u,v) = 0 \implies u = v$$

2. $d(x_n, x_{n+1}) \leq \alpha^n d(x_0, x_1)$:

$$\begin{aligned} d(x_n, x_{n+1}) &= d(F(x_{n-1}), F(x_n)) \leq \alpha d(x_{n-1}, x_n) \\ &\leq \alpha^2 d(x_{n-2}, x_{n-1}) \leq \ldots \leq \alpha^n d(x_0, x_1). \end{aligned}$$

3. $d(x_n, x_{n+p}) \leq \frac{\alpha^n}{1-\alpha}\, d(x_0, x_1)$ pro všechna $p > 0$: Užitím D3 a 2 lze ukázat

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \ldots + d(x_{n+p-1}, x_{n+p}) \\ &\leq (\alpha^n + \alpha^{n+1} + \ldots + \alpha^{n+p-1}) d(x_0, x_1) \\ &\leq \alpha^n (1 + \alpha + \alpha^2 + \ldots) d(x_0, x_1) = \frac{\alpha^n}{1-\alpha}\, d(x_0, x_1). \end{aligned}$$

4. EXISTENCE: Protože podle 3 platí $0 \leq d(x_n, x_{n+p}) \leq \frac{\alpha^n}{1-\alpha}\, d(x_0, x_1)$, $d(x_n, x_{n+p}) \longrightarrow 0$ pro $n \longrightarrow \infty$ a $p > 0$ libovolné. Tedy posloupnost $(x_n)_{n=0}^{\infty}$ je cauchyovská. Protože metrický prostor X je úplný, je tato posloupnost konvergentní a tedy existuje $\hat{x} = \lim_{n\to\infty} x_n$. Víme již, že potom $\hat{x}$ je řešením ÚLOHY.
5. $d(x_n, \hat{x}) \leq \alpha^n d(x_0, \hat{x})$:

$$d(x_n, \hat{x}) = d(F(x_{n-1}), F(\hat{x})) \leq \alpha d(x_{n-1}, \hat{x}) \leq \ldots \leq \alpha^n d(x_0, \hat{x})$$

6. $d(x_n, \hat{x}) \leq \frac{\alpha^n}{1-\alpha}\, d(x_0, x_1)$: Podle D3 a 3 platí

$$d(x_n, \hat{x}) \leq d(x_n, x_{n+p}) + d(x_{n+p}, \hat{x}) \longrightarrow \frac{\alpha^n}{1-\alpha}\, d(x_0, x_1) \quad \text{pro} \quad p \longrightarrow \infty.$$

Příklady metrických prostorů:

1. $E_1 = (R, d) \quad d(x,y) = |x - y|$
2. $E_2 = (R^2, d_2) \quad d_2(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

3. $E_n = (R^n, d_2) \quad d_2(x,y) = \sqrt{(x_1-y_1)^2 + \ldots + (x_n-y_n)^2}$
4. $(R^n, d_\infty) \quad d_\infty(x,y) = \max_{1\le i\le n} |x_i - y_i|$
5. $(R^n, d_1) \quad d_1(x,y) = \sum_{i=1}^n |x_i - y_i|$
6. $(C\langle a,b\rangle, d_\infty) \quad d_\infty(f,g) = \max_{a\le x\le b} |f(x) - g(x)|$
7. $(C\langle a,b\rangle, d_2) \quad d_2(f,g) = \sqrt{\int_a^b (f(x)-g(x))^2\,dx}$

Příklad 8. Metrický prostor (X,d), kde $X = (0,1\rangle$ a $d(x,y) = |x-y|$ není úplný:

Posloupnost $\left(\frac{1}{n}\right)_{n=1}^{\infty}$ v X není konvergentní, ale je cauchyovská:

$$\lim_{n\to\infty} \frac{1}{n} = 0 \implies \left(\frac{1}{n}\right) \text{ je konvergentní v } E_1$$
$$\implies \left(\frac{1}{n}\right) \text{ je cauchyovská v } E_1 \implies \left(\frac{1}{n}\right) \text{ je cauchyovská v } X.$$

Přitom tato posloupnost nemá v X limitu.

Poznámka. Metrický prostor (E_1, d) je úplný a prostory $(\langle a,b\rangle, d)$ jsou úplné pro všechny uzavřené intervaly $\langle a,b\rangle$. Prostory $E_n, n \ge 1$, (R^n, d_∞) a (R^n, d_1) jsou úplné. Prostor $(C\langle a,b\rangle, d_\infty)$ je úplný, ale prostor $(C\langle a,b\rangle, d_2)$ není úplný.

4 Iterační metody řešení jedné rovnice pro jednu reálnou neznámou

Metoda prosté iterace

$$\text{H}ledme \;\; x \in I: \qquad f(x) = 0, \tag{2}$$

kde I je libovolný interval v R.

Rovnice $f(x) = 0$ se převede na ekvivalentní tvar

$$x = F(x), \tag{3}$$

$x_0 \in I$ se zvolí a $x_{n+1} = F(x_n)$ pro $n = 0, 1, \ldots$ vytvoří iterační posloupnost, jejíž limita je řešením úlohy (2).

Věta 1. Nechť I je uzavřený interval v R a $F : I \to I$ zobrazení s vlastností $|F'(x)| \le \alpha < 1$ pro všechna $x \in I$. Pak F je kontrakce v úplném metrickém prostoru I s koeficientem α.

Důkaz. Stačí ověřit, že $d(F(x), F(y)) \leq \alpha d(x, y)$, tj. $|F(x) - F(y)| \leq \alpha |x - y|$ pro všechna $x, y \in I$: Podle Lagrangeovy věty o přírůstku funkce platí

$$F(x) - F(y) = F'(\xi)(x - y)$$

pro vhodný bod ξ mezi $x, y \in I$. Pak zřejmě $\xi \in I$ a tedy

$$|F(x) - F(y)| = |F'(\xi)|\,|x - y| \leq \alpha |x - y|.$$

Důsledek. Jsou-li splněny předpoklady Věty 1, pak pro řešení úlohy (3) platí všechna tvrzení Věty o kontrakci.

Příklad 1. Určete všechny kořeny rovnice $f(x) \equiv e^{-2x} + x - 3 = 0$ na čtyři desetinná místa.

Grafická metoda:

$$f(x) = 0 \Longleftrightarrow e^{-2x} = -x + 3.$$

Schematické znázornění grafů funkcí e^{-2x} a $-x + 3$ ilustruje skutečnost, že rovnice má právě dva kořeny $x^{(1)} \doteq -1$ a $x^{(2)} \doteq 2,8$. Za účelem řešení rovnice iterací převedeme původní rovnici $f(x) = 0$ na tvar (3). a) $f(x) = 0 \Leftrightarrow x = 3 - e^{-2x} \equiv F_1(x)$. Protože

$$|F_1'(x)| = 2e^{-2x} < 1 \Leftrightarrow x > \frac{1}{2} \ln 2 \doteq 0,3466,$$

užijeme této ekvivalentní formulace pro aproximaci kořene $x^{(2)}$. Položíme $x_0 = 2,8$ a pro $n = 0, 1, \ldots$ budeme postupně počítat

$$x_{n+1} = 3 - e^{-2x_n}.$$

Viz tabulku níže.

n	x_n
0	2,8
1	2,9963
2	2,9975
3	2,9975

b) $f(x) = 0 \Longleftrightarrow x = -0,5 \ln(3 - x) \equiv F_2(x)$. Protože

$$|F_2'(x)| = \frac{1}{2(3 - x)} < 1 \Longleftrightarrow x < 2,5,$$

lze této formulace užít pro aproximaci kořene $x^{(1)}$ podle předpisu $x_0 = -1$ a

$$x_{n+1} = -0,5 \ln(3 - x_n)$$

pro $n = 0, 1, \ldots$. Viz tabulku.

n	x_n
0	-1
1	-0,6931
2	-0,6532
3	-0,6478
4	-0,6471
5	-0,6470
6	-0,6469
7	-0,6469

Newtonova metoda

Předpokládejme, že aproximace x_n leží blízko kořene $\hat{x}$ rovnice (2). Pak

$$0 = f(\hat{x}) = f(x_n) + f'(x_n)(\hat{x} - x_n) + \frac{f''(\xi)}{2}(\hat{x} - x_n)^2 \tag{4}$$

Předpokládejme, že $f'(x_n) \neq 0$ a vydělme (3) $f'(x_n)$. Osamostatněním $\hat{x}$ vznikne

$$\hat{x} = x_n - \frac{f(x_n)}{f'(x_n)} - K(\hat{x} - x_n)^2 \quad \text{pro} \quad K = \frac{f''(\xi)}{2f'(x_n)} \tag{5}$$

V (4) zanedbáme poslední člen a $\hat{x}$ nahradíme hodnotou x_{n+1}. Vznikne tento *předpis Newtonovy metody:*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{6}$$

Příklad 2.

$$f(x) \equiv e^{-2x} + x - 3 = 0, \qquad f'(x) = 1 - 2e^{-2x}$$
$$x_0 = -1, \quad x_{n+1} = x_n - \frac{e^{-2x_n} + x_n - 3}{1 - 2e^{-2x_n}} \text{ pro } \quad n = 0, 1, \ldots$$

n	x_n
0	-1
1	-0,7540
2	-0,6589650106
3	-0,6471103830
4	-0,6469449337
5	-0,6469449022
6	-0,6469449022

Poznámka. Odečtením (5) od (4) vznikne

$$|\hat{x} - x_{n+1}| = |K|(\hat{x} - x_n)^2, \quad tj. \quad |e_{n+1}| \leq |K||e_n|^2$$

O metodě půlení intervalu víme, že pro odhad ε_k chyby $e_k = \hat{x} - s_k$ platí

$$\varepsilon_{n+1} \leq \frac{1}{2}\varepsilon_n$$

a pro metodu prosté iterace platí

$$|e_{n+1}| \leq \alpha |e_n|,$$

kde α je koeficient kontrakce.

Definice. Řekneme, že iterační metoda je řádu r, jestliže odhady chyb splňují nerovnost

$$\varepsilon_{n+1} \leq C\varepsilon_n^r$$

pro všechna n a pro C nezávislé na n.

Tedy Newtonova metoda je řádu 2 a metoda půlení intervalu i metoda prosté iterace jsou řádu 1. Metoda regula falsi je rovněž řádu 1.

5 Vektorové prostory

Libovolný vektor $\vec{x} \in R^n$ budeme považovat za sloupcový a budeme jej podrobněji značit

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^\top$$

Vzhledem k dobře známým vlastnostem nulového vektoru $\vec{o}$ a operací sčítání vektorů a násobení vektorů reálnými čísly lze vytvářet lineární kombinace vektorů.

Definice. Pro libovolná čísla $c_1, c_2, \dots, c_k \in R$ a vektory $\vec{x}^1, \vec{x}^2, \dots, \vec{x}^k \in R^n$ se výraz

$$c_1\vec{x}^1 + c_2\vec{x}^2 + \dots + c_k\vec{x}^k$$

nazývá *lineární kombinace* vektorů $\vec{x}^1, \vec{x}^2, \dots, \vec{x}^k$ s koeficienty $c_1, c_2, \dots, c_k$.

Definice. Množinu všech lineárních kombinací vektorů $\vec{x}^1, \vec{x}^2, \dots, \vec{x}^k$ budeme značit

$$\mathcal{L}(\vec{x}^1, \vec{x}^2, \dots, \vec{x}^k).$$

Definice. Vektory $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$ se nazývají *lineárně nezávislé*, jestliže platí

$$c_1\vec{x}^1 + c_2\vec{x}^2 + \ldots + c_k\vec{x}^k = \vec{o} \tag{7}$$

jen tehdy, když $c_1 = c_2 = \ldots = c_k = 0$.

Existují-li $c_1, c_2, \ldots, c_k$, ne všechna rovna nule, splňující (7), pak se vektory $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$ nazývají *lineárně závislé*.

Příklad 1. Vektory $\vec{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\vec{x}_2 = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$ jsou lineárně závislé:

$$c_1\vec{x}^1 + c_2\vec{x}^2 = \begin{bmatrix} c_1 - 2c_2 \\ 2c_1 - 4c_2 \end{bmatrix} = \vec{o} \Longleftrightarrow \begin{array}{c} c_1 - 2c_2 = 0 \\ 2c_1 - 4c_2 = 0 \end{array} \Longleftrightarrow c_1 = 2c_2.$$

Vektory $\vec{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\vec{x}_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$ jsou lineárně nezávislé:

$$c_1\vec{x}^1 + c_2\vec{x}^2 = \begin{bmatrix} c_1 - 2c_2 \\ 2c_1 \end{bmatrix} = \vec{o} \Longleftrightarrow \begin{array}{c} c_1 - 2c_2 = 0 \\ 2c_1 = 0 \end{array} \Longleftrightarrow c_1 = 0 = c_2.$$

Věta 1. Vektory $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$ jsou lineárně závislé právě když je jeden z nich lineární kombinací ostatních.

Důkaz. $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$ lineárně závislé $\Longleftrightarrow$ existují $c_1, c_2, \ldots, c_k$, ne všechna rovna nule tak, že $c_1\vec{x}^1 + \ldots + c_k\vec{x}^k = \vec{o}$ (nechť například $c_1 \neq 0$) $\Longleftrightarrow \vec{x}^1 = -\frac{c_2}{c_1}\vec{x}^2 - \ldots - \frac{c_k}{c_1}\vec{x}^k$.

Definice. Neprázdná množina $V \subset R^n$ se nazývá *vektorový prostor*, jestliže

$$\begin{array}{rcl} \vec{x} \in V, a \in R & \Longrightarrow & a\vec{x} \in V \\ \vec{x} \in V, \vec{y} \in V & \Longrightarrow & \vec{x} + \vec{y} \in V. \end{array}$$

Příklad 2. Množina $\mathcal{L}(\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k)$ je vektorový prostor pro libovolné vektory $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$.

Důkaz cvičení.

Definice. Vektory $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$ z vektorového prostoru V tvoří bazi ve V, jestliže platí

a) $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k$ jsou lineárně nezávislé

b) $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k, \vec{x}$ jsou lineárně závislé pro všechna $\vec{x} \in V$.

Dimenze vektorového prostoru je počet vektorů v jeho libovolné bazi.

Definice. Podmnožina W vektorového prostoru V se nazývá *podprostor* ve V, je-li W vektorový prostor.

Poznámka. Libovolnou matici A typu (n, k) lze považovat za vektor s $n \cdot k$ složkami. Jsou-li A, B matice téhož typu a $\alpha \in R$, pak

$$\alpha A = (\alpha a_{ij}) \quad \text{a} \quad A + B = (a_{ij} + b_{ij})$$

Tedy množina všech matic téhož typu je vektorový prostor.

Definice. Čtvercová matice A se nazývá $\left\{ \begin{array}{l} \text{horní trojúhelníková} \\ \text{dolní trojúhelníková} \\ \text{diagonální} \\ \text{symetrická} \end{array} \right\}$, když $\left\{ \begin{array}{l} a_{ij} = 0 \ \text{ pro všechna } \ i > j \\ a_{ij} = 0 \ \text{ pro všechna } \ i < j \\ a_{ij} = 0 \ \text{ pro všechna } \ i \neq j \\ a_{ij} = a_{ji} \ \text{ pro všechna } \ i, j \end{array} \right\}$. Symbol $\operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$ značí diagonální matici A.

Příklad 3. Množina všech $\left\{ \begin{array}{l} \text{horních trojúhelníkových} \\ \text{dolních trojúhelníkových} \\ \text{diagonálních} \\ \text{symetrických} \end{array} \right\}$ matic řádu n tvoří vektorový prostor. Je to podprostor v prostoru všech matic řádu n.

Definice. Jednotkovou matici E a nulovou matici O lze definovat předpisem

$$E = \operatorname{diag}(1, 1, \ldots, 1) \quad \text{a} \quad O = \operatorname{diag}(0, 0, \ldots, 0).$$

Definice. Čtvercová matice A se nazývá *regulární*, když $\det A \neq 0$. Jinak se A nazývá *singulární*.

Definice. Jsou-li matice A typu (m, n) a matice B typu (n, k), pak matice $C = A \cdot B$ je typu (m, k) a pro její prvky platí $c_{ij} = a_{i1}b_{1j} + \ldots + a_{in}b_{nj}$ pro všechna $i, j = 1, \ldots, n$.

Poznámka. Pro libovolnou čtvercovou matici A platí $A + O = A$ a $AE = EA = A$. Pro libovolný vektor $\vec{x}$ platí $E\vec{x} = \vec{x}$.

Věta 2.

a) Ke každé regulární matici A existuje jediná matice A^{-1} tak, že $A\,A^{-1} = A^{-1}\,A = E$. Pak se matice A^{-1} nazývá *inverzní* k matici A.

b) Jsou-li A, B regulární matice téhož řádu, pak AB je regulární a platí

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Definice. Matici A typu (m, n) lze považovat za soubor vektorů $\vec{a}^1, \ldots, \vec{a}^n$ z R^m (sloupcových vektorů A) a značit $A = [\vec{a}^1, \ldots, \vec{a}^n]$. Dimenze vektorového prostoru $\mathcal{L}(\vec{a}^1, \ldots, \vec{a}^n)$ se nazývá *hodnost* matice A a značí se $h(A)$.

6 Normy matic a vektorů

Definice. Předpis, který k libovolné matici A přiřadí reálné číslo $\|A\|$ se nazývá *norma matic*, platí-li

N1 $\|A\| \geq 0$ a $\|A\| = 0 \Longleftrightarrow A = 0$

N2 $\|\alpha A\| = |\alpha| \|A\|$

N3 $\|A + B\| \leq \|A\| + \|B\|$

N4 $\|A \cdot B\| \leq \|A\| \cdot \|B\|$

Věta 1. Každý předpis, který k libovolné matici A typu (m, n) přiřadí číslo

a) $\|A\|_r = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$,

b) $\|A\|_s = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|$,

c) $\|A\|_e = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$

je norma. Tyto předpisy se postupně nazývají *řádková*, *sloupcová* a *euklidovská norma*.

Poznámka. Libovolný vektor $\vec{x} \in R^n$ je matice typu $(n, 1)$. Každé zobrazení $x \longmapsto \|x\|$ z R^n do R s vlastnostmi N1, N2, N3 se nazývá *norma vektorů*. Podle Věty 1 jsou předpisy

$$\|\vec{x}\|_r = \max_{1 \leq i \leq n} |x_i|, \quad \|\vec{x}\|_s = \sum_{i=1}^{n} |x_i|, \quad \|\vec{x}\|_e = \sqrt{\sum_{i=1}^{n} x_i^2}$$

normy vektorů.

Věta 2. Nechť A je řádu n a $\vec{x} \in R^n$. Pak pro $Q = r, s, e$ platí

$$\|A\vec{x}\|_Q \leq \|A\|_Q \|\vec{x}\|_Q.$$

Tato nerovnost se nazývá *podmínka konzistence*.

Definice. Skalární součin vektorů $\vec{x}, \vec{y} \in R^n$ je číslo

$$(\vec{x}, \vec{y}) = \vec{x}^\top \vec{y} = x_1 y_1 + \ldots + x_n y_n.$$

Věta 3. Skalární součin má vlastnosti

S1 $(\vec{x},\vec{x}) \geq 0$ a $(\vec{x},\vec{x}) = 0 \Longleftrightarrow \vec{x} = \vec{o}$

S2 $(\vec{x},\vec{y}) = (\vec{y},\vec{x})$

S3 $(\alpha\vec{x},\vec{y}) = \alpha(\vec{y},\vec{x})$

S4 $(\vec{x},\vec{y}+\vec{z}) = (\vec{x},\vec{y}) + (\vec{x},\vec{z})$

a platí:

$$(\vec{x},\vec{x}) = \|x\|_e^2, \quad |(\vec{x},\vec{y})| \leq \|x\|_e \cdot \|y\|_e \quad \text{(Cauchyova nerovnost)}$$

Důkaz Cauchyovy nerovnosti: Pro všechna reálná čísla λ platí

$$\begin{array}{rcl} 0 & \leq & (\vec{x}+\lambda\vec{y},\vec{x}+\lambda\vec{y}) = (\vec{x},\vec{x}) + 2\lambda(\vec{x},\vec{y}) + \lambda^2(\vec{y},\vec{y}) \\ & \Longleftrightarrow & 4(\vec{x},\vec{y})^2 - 4(\vec{x},\vec{x})(\vec{y},\vec{y}) \leq 0 \Longleftrightarrow (\vec{x},\vec{y})^2 \leq (\vec{x},\vec{x})(\vec{y},\vec{y}) \\ & \Longleftrightarrow & |(\vec{x},\vec{y})| \leq \|x\|_e \cdot \|y\|_e \end{array}$$

7 Finitní metody řešení systémů lineárních algebraických rovnic

$$A\vec{x} = \vec{b} \tag{8}$$

Předpoklad: Matice A je regulární (a tedy i čtvercová).

Poznámka. Úloha (8) má jediné řešení $\hat{x}$. Numerický výpočet poskytne vždy jen aproximaci $\vec{x}$ vektoru $\hat{x}$. Numerické metody řešení úlohy (8) jsou v podstatě dvojího typu:

a) finitní (přímé), které (teoreticky) přesným provedením předepsaného konečného počtu operací poskytnou přesné řešení $\hat{x}$.

b) iterační.

Gaussova eliminační metoda (GEM)

je založena na tom, že řešení soustavy (9), v níž $u_{ii} \neq 0$ pro $i = 1,\ldots,n$, je velmi rychlé.

$$\begin{array}{rcl} u_{11}x_1 + u_{12}x_2 + \ldots + u_{1n}x_n & = & d_1 \\ \vdots & & \\ u_{n-1n-1}x_{n-1} + u_{n-1n}x_n & = & d_{n-1} \\ u_{nn}x_n & = & d_n \end{array} \tag{9}$$

Poznámka. Lze snadno ověřit, že počet operací násobení a dělení (i sčítání a odčítání) při řešení tohoto systému n rovnic s horní trojúhelníkovou maticí je přibližně $\frac{n^2}{2}$.

GEM má dvě části:

1. Přímý chod. Převedení dané soustavy na soustavu s horní trojúhelníkovou maticí opakováním ekvivalentní úpravy "přičtení násobku jedné rovnice k rovnici jiné".

2. Zpětný chod. Řešení soustavy s horní trojúhelníkovou maticí.

Příklad 1. Převeďte daný systém rovnic na soustavu s horní trojúhelníkovou maticí.

$$
\begin{array}{rrcll}
 & \boxed{1}x_1 + 4x_2 + 3x_3 & = & 1 & \\
\text{1. fáze} & 2x_1 + 5x_2 + 4x_3 & = & 4 & m_{21} = -2 \\
 & x_1 - 3x_2 - 2x_3 & = & 5 & m_{31} = -1 \\
\hline
 & \boxed{\text{-3}}x_2 - 2x_3 & = & 2 & \\
\text{2. fáze} & - 7x_2 - 5x_3 & = & 4 & m_{32} = -\dfrac{7}{3} \\
\hline
 & \boxed{-\frac{1}{3}}x_3 & = & -\dfrac{2}{3} &
\end{array}
$$

Výsledný systém rovnic s horní trojúhelníkovou maticí tvoří právě rovnice, v nichž jsou koeficienty v rámečku.

Poznámka. Počet násobení a dělení i počet sčítání a odčítání v přímém chodu GEM je až na násobky nižších mocnin roven $\frac{n^3}{3}$.

Definice. Hlavní prvek (k-té fáze) je prvek, který je ve jmenovateli multiplikátorů z k-té fáze a kterým se dělí ve zpětném chodu při výpočtu x_k.

Definice. Nechť A je matice řádu n a $k \in \{1, 2, \ldots, n\}$. Označme $A^{(k)}$ matici, vytvořenou z průsečíků prvních k řádků a k sloupců matice A. Tedy

$$A^{(1)} = [a_{11}], \quad A^{(2)} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \ldots \quad , A^{(n)} = A.$$

Věta 1. Systém rovnic $A\vec{x} = \vec{b}$ je řešitelný GEM právě tehdy, když $\det A^{(k)} \neq 0$ pro $k = 1, \ldots, n$.

Příklad 2. Níže uvedený systém rovnic řešte GEM. Zaokrouhlujte na 4 platné číslice.

$$
\begin{array}{rcll}
x_1 + x_2 + x_3 & = & 1 & \\
0.0001x_2 + x_3 & = & 1 & m_{32} = -\dfrac{1}{0.0001} = -10000 \\
x_2 + x_3 & = & 0 & \\
\hline
-9999x_3 & = & -10000 &
\end{array}
$$

Tedy $\hat{x}_3 = \frac{10000}{9999} = 1.000\overline{1000} \Longrightarrow x_3 = 1$ a odtud plyne

$$\hat{x} = \begin{bmatrix} 1 \\ -\frac{10000}{9999} \\ \frac{10000}{9999} \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Zdůvodnění podstatného rozdílu mezi $\hat{x}$ a $\vec{x}$: Při výpočtu x_3 vznikla zaokrouhlovací chyba $\frac{1}{9999}$. Při výpočtu x_2 se tato chyba zvětšila 10000 krát na $\frac{10000}{9999}$. Důvodem tohoto nárůstu je fakt, že hodnota hlavního prvku druhé fáze je $0.0001 \ll 1$.

GEM s částečným a úplným výběrem hlavních prvků

A) Částečný výběr: V k-té fázi se za hlavní prvek vybere v absolutní hodnotě největší číslo z k-tého sloupce, ležící v nebo pod hlavní diagonálou.

B) Úplný výběr: V k-té fázi se za hlavní prvek vybere v absolutní hodnotě největší číslo s řádkovým i sloupcovým indexem mezi k a n.

7.1 LU-rozklad matice

Definice. Nechť A je čtvercová matice řádu n. Dolní trojúhelníková matice L s jednotkami v hlavní diagonále a horní trojúhelníková matice U tvoří LU-rozklad matice A, jestliže $A = LU$.

Příklad 3. Nechť $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, $M_1 = \begin{bmatrix} 1 & & \\ m_{21} & 1 & \\ m_{31} & & 1 \end{bmatrix}$

a $M_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & m_{32} & 1 \end{bmatrix}$. Pak $M_1 A =$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ m_{21}a_{11} + a_{21} & m_{21}a_{12} + a_{22} & m_{21}a_{13} + a_{23} \\ m_{31}a_{11} + a_{31} & m_{31}a_{12} + a_{32} & m_{31}a_{13} + a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ & a_{22}^{(1)} & a_{23}^{(1)} \\ & a_{32}^{(1)} & a_{33}^{(1)} \end{bmatrix} \text{ a}$$

$$M_2(M_1A) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ & a_{22}^{(1)} & a_{23}^{(1)} \\ & m_{32}a_{22}^{(1)} + a_{32}^{(1)} & m_{32}a_{23}^{(1)}a_{33}^{(1)} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ & a_{22}^{(1)} & a_{23}^{(1)} \\ & & a_{33}^{(2)} \end{bmatrix} = U.$$

Horní trojúhelníková matice $M_2(M_1A) = U$ je výsledkem přímého chodu GEM. Násobíme-li obě strany této maticové rovnice zleva postupně maticemi M_2^{-1} a M_1^{-1}, vznikne

$$A = (M_1^{-1}M_2^{-1}) = LU.$$

Lze snadno ověřit, že

$$M_1^{-1} = \begin{bmatrix} 1 & & \\ -m_{21} & 1 & \\ -m_{31} & & 1 \end{bmatrix}, \quad M_2^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & -m_{32} & 1 \end{bmatrix}$$

$$\text{a tedy} \quad M_2^{-1}M_1^{-1} = \begin{bmatrix} 1 & & \\ -m_{21} & 1 & \\ -m_{31} & -m_{32} & 1 \end{bmatrix} = L$$

je horní trojúhelníková matice, takže L, U tvoří LU-rozklad matice A. Je zřejmé, že tento rozklad je produktem přímého chodu GEM. Matice L je sestavena z multiplikátorů přímého chodu a matice U je výsledkem přímého chodu.

Je zřejmé, že stejná tvrzení platí i pro matice jiných řádů, než 3.

Příklad 4. Matice soustavy rovnic z příkladu 1 má tento LU-rozklad:

$$\begin{bmatrix} 1 & 4 & 3 \\ 2 & 5 & 4 \\ 1 & -3 & -2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ 1 & \frac{7}{3} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 4 & 3 \\ & -3 & 2 \\ & & -\frac{1}{3} \end{bmatrix}.$$

Věta 2. Nechť A je matice řádu n splňující $\det A^{(k)} \neq 0$ pro $k = 1, \ldots, n-1$. Pak $A = LU$, kde

$$L = \begin{bmatrix} 1 & & & \\ -m_{21} & 1 & & \\ \vdots & \vdots & \ldots & \\ -m_{n1} & -m_{n2} & \ldots & 1 \end{bmatrix}, \quad U = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ & a_{22}^{(1)} & \ldots & a_{2n}^{(1)} \\ & & \ldots & \vdots \\ & & & a_{nn}^{(n-1)} \end{bmatrix}.$$

7.2 Výpočet matice inverzní

Předpokládejme, že matice A je regulární řádu n. Je dobře známo, že matice X je inverzní k A, když

$$AX = E. \tag{10}$$

Označíme-li matice X, E pomocí sloupcových vektorů, tj. $X = [\vec{x}^1, \ldots, \vec{x}^n]$, $E = [\vec{e}^1, \ldots, \vec{e}^n]$, pak (10) $\Longleftrightarrow$ (11), kde

$$A\vec{x}^j = \vec{e}^j \quad \text{pro} \quad j = 1, \ldots, n, \tag{11}$$

což je soustava n systémů rovnic se stejnou maticí A. Všechny systémy rovnic této soustavy lze řešit současně.

Příklad 5. Určete matici, inverzní k $A = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 5 & 4 \\ 1 & -3 & -2 \end{bmatrix}$.

$\boxed{1}$	4	3	1	0	0
2	5	4	0	1	0
1	-3	-2	0	0	1
	$\boxed{-3}$	-2	-2	1	0
	-7	-5	-1	0	1
		$\boxed{-\frac{1}{3}}$	$\frac{11}{3}$	$-\frac{7}{3}$	1

Použitím tří zpětných chodů získáme sloupcové vektory

$$\vec{x}^1 = \begin{bmatrix} 2 \\ 8 \\ -11 \end{bmatrix}, \quad \vec{x}^2 = \begin{bmatrix} -1 \\ -5 \\ 7 \end{bmatrix}, \quad \vec{x}^3 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \quad \text{a tedy}$$

$$A^{-1} = \begin{bmatrix} 2 & -1 & 1 \\ 8 & -5 & 2 \\ -11 & 7 & -3 \end{bmatrix}.$$

Poznámka. Lze ověřit, že algoritmus výpočtu inverzní matice potřebuje zhruba n^3 operací násobení a dělení a zhruba stejný počet operací sčítání a odčítání.

7.3 Číslo podmíněnosti matice

Nyní se budeme zabývat otázkou závislosti chyby řešení systému rovnic (8), tj. $A\hat{x} = \hat{b}$, na chybě vektoru $\vec{b}$ pravých stran za teoretického předpokladu, že při řešení jsou všechny aritmetické operace prováděny přesně. Označíme

$\vec{b}$ aproximaci přesného řešení $\quad \hat{b} = \vec{b} + \vec{e}_b$

$\vec{x}$ přesné řešení soustavy (8)

$\hat{x}$ přesné řešení soustavy $\mathrm{A\hat{x} = \hat{b}}$ $\quad \hat{x} = \vec{x} + \vec{e}_x$

Symbolem $\|\cdot\|$ bude v následující úvaze značena libovolná norma matic a s ní konzistentní norma vektorů.

$$\begin{aligned} A(\vec{x} + \vec{e}_x) &= \vec{b} + \vec{e}_x \Longrightarrow \\ A\vec{x} + A\vec{e}_x &= \vec{b} + \vec{e}_x \Longrightarrow \quad \text{podle (8)} \\ A\vec{e}_x = \vec{e}_b &\Longrightarrow \vec{e}_x = A^{-1}\vec{e}_b \Longrightarrow \\ \|\vec{e}_x\| &\leq \|A^{-1}\|\,\|\vec{e}_b\|. \end{aligned}$$

Tato nerovnost a nerovnost $\|\vec{b}\| \leq \|A\|\,\|\vec{x}\|$, která je důsledkem (8), implikují

$$\|\vec{e}_x\|\,\|\vec{b}\| \leq \|A\|\,\|A^{-1}\|\,\|\vec{e}_b\|\|\vec{x}\|$$

a odtud po vydělení obou stran výrazem $\|\vec{b}\|\,\|\vec{x}\|$ vznikne nerovnost

$$\frac{\|\vec{e}_x\|}{\|\vec{x}\|} \leq \|A\|\,\|A^{-1}\|\frac{\|\vec{e}_b\|}{\|\vec{b}\|}. \tag{12}$$

Je zřejmé, že levá strana této nerovnosti odpovídá relativní chybě vektoru řešení a pravá strana odpovídá relativní chybě vektoru pravých stran, vynásobené koeficientem

$$\mathrm{cond}A = \|A\|\,\|A^{-1}\|,$$

který se nazývá *číslo podmíněnosti matice A*. Je-li číslo podmíněnosti matice A velké [malé], nazývá se A dobře [špatně] podmíněná.

Příklad 1. Pro systém rovnic

$$\begin{aligned} x_1 + 0,7x_2 &= 1,69 \\ 0,7x_1 + 0,5x_2 &= 1,21 \end{aligned}$$

označme $A = \begin{bmatrix} 1 & 0,7 \\ 0,7 & 0,5 \end{bmatrix}$, $\hat{b} = \begin{bmatrix} 1,69 \\ 1,21 \end{bmatrix}$ a položme $\vec{b} = \begin{bmatrix} 1,7 \\ 1,2 \end{bmatrix}$. Současným řešením systémů rovnic $A\hat{x} = \hat{b}$ a $A\vec{x} = \vec{b}$ lze snadno zjistit, že

$$\hat{x} = \begin{bmatrix} -0.2 \\ 2.7 \end{bmatrix} \quad \text{a} \quad \vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{a tedy} \quad \vec{e}_x = \begin{bmatrix} -1,2 \\ 1,7 \end{bmatrix}.$$

Protože $A^{-1} = \begin{bmatrix} 50 & -70 \\ -70 & 100 \end{bmatrix}$, platí pro řádkovou normu matic a vektorů

$$\|\vec{e}_x\| = 1,7, \quad \|\vec{x}\| = 1, \quad \|A\| = 1,7, \quad \|A^{-1}\| = 170, \quad \|\vec{e}_b\| = 0,01 \text{ a } \|\vec{b}\| = 1,7.$$

Dosazením těchto hodnot do nerovnosti (12) vznikne

$$\frac{1,7}{1} \leq 170 \cdot 1,7 \frac{0,01}{1,7},$$

takže v tomto případě je nerovnost (12) splněna rovností. Příklad ukazuje, že horní odhad relativní chyby vektoru řeešení v (12) není zbytečně příliš velký.

7.4 Speciální matice soustavy

(A) Pozitivně definitní matice

Definice. Matice A řádu n se nazývá *pozitivně definitní*, je-li

a) symetrická a

b) platí-li

$$\vec{x}^\top A\vec{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j > 0 \quad \text{pro všechna} \quad \vec{x} \neq \vec{o}.$$

Věta 3. Symetrická matice A řádu n je pozitivně definitní právě když jsou všechny hlavní prvky $a_{11}, a_{22}^{(1)}, \ldots, a_{nn}^{(n-1)}$ z GEM kladné.

Věta 4. Je-li matice A symetrická, pak je symetrická i matice

$$A_k = \begin{bmatrix} a_{k+1,k+1}^{(k)} & \cdots & a_{k+1,n}^{(k)} \\ \vdots & & \vdots \\ a_{n,k+1}^{(k)} & \cdots & a_{n,n}^{(k)} \end{bmatrix}$$

pro $k = 1, \ldots, n-1$.

Důsledek. Je-li v úloze (8) matice A pozitivně definitní, lze (8) řešit GEM a v přímém chodu počítat jen prvky v a nad hlavní diagonálou. Čas i nároky na operační paměť se redukují v podstatě na polovinu.

Věta 5. Je-li matice A pozitivně definitní, pak existuje jediná horní trojúhelníková matice U s kladnými prvky v hlavní diagonále taková, že $A = U^\top U$.

Poznámka. Je-li známá matice U z Věty 5, lze místo soustavy rovnic $A\vec{x} = \vec{b}$ řešit dvě soustavy $U^\top \vec{y} = \vec{b}$ a potom $U\vec{x} = \vec{y}$. Algoritmus pro získání matice

$$U = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ & u_{22} & \dots & u_{2n} \\ & & \dots & \vdots \\ & & & u_{nn} \end{bmatrix}$$

je touto modifikací přímého chodu GEM:

1. $u_{11} = \sqrt{a_{11}}$ a pro $j = 2, \dots, n$:
$$u_{1j} = a_{1j}/u_{11}$$
2. Pro $i = 2, \dots, n-1$: $u_{ii} = \sqrt{a_{ii} - \sum_{r=1}^{i-1} u_{ri}^2}$ a pro $j = i+1, \dots, n$:
$$u_{ij} = \left(a_{ij} - \sum_{s=1}^{i-1} u_{si} u_{sj} \right) / u_{ii}$$
3. $u_{nn} = \sqrt{a_{nn} - \sum_{r=1}^{n-1} u_{rn}^2}$

Tato metoda se nazývá *Choleského metoda* nebo *odmocninová metoda* a je velmi populární. Je časově stejně náročná jako metoda využívající vět 3, 4.

(B) Pásové matice

Definice. Matice A řádu n se nazývá *pásová*, existují-li $p, q \geq 0$:

$$j > i + p \Longrightarrow a_{ij} = 0 \quad \text{a} \quad i > j + q \Longrightarrow a_{ij} = 0 \quad \text{pro} \quad i, j = 1, \dots, n.$$

Číslo $s = p + q + 1$ se nazývá *šířka pásu* matice A. Při LU-rozkladu matice A se pásovost zachovává.

Poznámka. Důležitou roli hrají tzv. *třídiagonální matice*. Jsou to pásové matice, pro něž $p = 1 = q$ a jejich LU-rozklad má tvar

$$A = \begin{bmatrix} a_1 & c_1 & & & & \\ b_2 & a_2 & c_2 & & & \\ & \dots & & & & \\ & & b_{n-1} & a_{n-1} & c_{n-1} & \\ & & & b_n & a_n & \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & & & & \\ \beta_2 & 1 & & & \\ & \dots & & & \\ & & \beta_{n-1} & 1 & \\ & & & \beta_n & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 & c_1 & & & \\ & \alpha_2 & c_2 & & \\ & \dots & & & \\ & & & \alpha_{n-1} & c_{n-1} \\ & & & & \alpha_n \end{bmatrix} = LU.$$

Porovnáním prvků matice A a součinu LU vzniknou tyto rovnice:

$$a_1 = \alpha_1 \quad \text{a} \quad b_k = \beta_k \alpha_{k-1}, \quad a_k = \beta_k c_{k-1} + \alpha_k \quad \text{pro} \quad k = 2, \ldots, n.$$

Užitím těchto rovnic lze spočítat prvky matic L a U tímto algoritmem:

$\alpha_1 := a_1;$

for $k := 2$ to n do

begin

$\beta_k = b_k / \alpha_{k-1};$

$\alpha_k = a_k - \beta_k c_{k-1}$

end;

Systém rovnic $A\vec{x} = \vec{d}$ lze pak řešit pomocí dvou zpětných chodů

$$L\vec{y} = \vec{d} \quad \text{a potom} \quad U\vec{x} = \vec{y}.$$

Tento algoritmus řešení lze pomocá fragmentu programu v jazyku Pascal zapsat takto:

$y_1 := d_1;$

for $k := 2$ to n do

$y_k = d_k - \beta_k y_{k-1};$

for $k := n - 1$ downto 1 do

$x_k = (y_k - c_k x_{k+1}) / \alpha_k;$

Z uvedených popisů algoritmů je patrné, že pro řešení systémů n rovnic pro n neznámých s třídiagonální maticí je třeba $5n - 5$ (přibližně $5n$) operací násobení a dělení a $3n - 3$ (přibližně $3n$) operací sčítání a odčítání. Porovnání těchto počtů s číslem $\frac{n^3}{3}$ vede k závěru, že řešení systémů rovnic s třídiagonálními maticemi je velmi efektivní.

Definice. Čtvercová matice se nazývá *řídká*, je-li většina jejích prvků rovna nule.

8 Vlastní čísla a vlastní vektory matic

Definice. Nechť A je matice řádu n. Jestliže pro číslo λ (obecně komplexní) a vektor $\vec{u} \neq \vec{o}$ platí

$$A\vec{u} = \lambda\vec{u}, \tag{13}$$

pak se λ nazývá *vlastní číslo* matice A a vektor $\vec{u}$ se nazývá *vlastní vektor* matice A, příslušný vlastnímu číslu λ.

Poznámka. Rovnici (13) lze psát ve tvaru

$$(A - \lambda E)\vec{u} = \vec{o}, \tag{14}$$

což je homogenní soustava n lineárních rovnic pro n neznámých. (14) má nenulové řešení právě když

$$\det(A - \lambda E) = 0. \tag{15}$$

Tato *charakteristická rovnice* matice A je polynomem n-tého stupně v λ. Tedy esistuje právě n (reálných či komplexních, případně násobných) vlastních čísel A. Pak

$$\varrho(A) = \max\{|\lambda|; \lambda \ \text{je } \textit{vlastn slo matice} \ \ A\} \tag{16}$$

se nazývá *spektrální poloměr* matice A.

Poznámka. Je-li známé vlastní číslo λ, pak příslušný vlastní vektor je každé nenulové řešení soustavy (14). Naopak, je-li známý vektor $\vec{u}$, pak z (13) plyne $\vec{u}^\top A\vec{u} = \lambda\vec{u}^\top\vec{u} = \lambda\|\vec{u}\|_E^2$. Odtud plyne, že

$$\lambda = \frac{\vec{u}^\top A\vec{u}}{\|\vec{u}\|_E^2}.$$

Toto vyjádření vlastního čísla λ se nazývá *Rayleighův podíl*.

Věta 1. Jestliže $A\vec{u} = \lambda\vec{u}$ pro $\vec{u} \neq 0$, $c \neq 0$ je reálné číslo a $k = 2, 3, \ldots$, pak platí

(a) $A(c\vec{u}) = \lambda(c\vec{u})$

(b) $(A - cE)\vec{u} = (\lambda - c)\vec{u}$

(c) $A^k\vec{u} = \lambda^k\vec{u}$

(d) $A^{-1}\vec{u} = \frac{1}{\lambda}\vec{u}$ pro regulární matici A.

Věta 2. Nechť matice A je symetrická. Pak

(a) všechna vlastní čísla matice A jsou reálná,

(b) vlastní vektory matice A, příslušné vzájemně různým vlastním číslům, jsou vzájemně kolmé a

(c) je-li A norma matic, konzistentní s odpovídající normou vektorů, pak $\varrho(A) \leq \|A\|$.

Důkaz (b): Nechť $\lambda \neq \mu$, $\vec{u}$ je vlastní vektor příslušný λ a $\vec{v}$ je vlastní vektor příslušný μ. Pak platí $A\vec{u} = \lambda\vec{u}$ a $A\vec{v} = \mu\vec{v}$ a odtud plyne

$$\begin{aligned}
\vec{v}^\top A\vec{u} &= \lambda\vec{v}^\top\vec{u} \quad |^\top \\
\vec{u}^\top A^\top\vec{v} &= \lambda\vec{u}^\top\vec{v} \\
\vec{u}^\top A\vec{v} &= \lambda\vec{u}^\top\vec{v} \\
\vec{u}^\top\mu\vec{v} &= \lambda\vec{u}^\top\vec{v} \\
\mu\vec{u}^\top\vec{v} &= \lambda\vec{u}^\top\vec{v} \\
\vec{u}^\top \cdot \vec{v} &= 0, \quad \text{nebo} \quad \lambda \neq \mu
\end{aligned}$$

Důkaz (c): Zvolme vlastní číslo λ tak, aby $|\lambda| = \varrho(A)$. Pak

$$\|A\|\,\|\vec{u}\| \geq \|A\vec{u}\| = \|\lambda\vec{u}\| = |\lambda|\|\vec{u}\| = \varrho(A)\|\vec{u}\|.$$

Definice. Soustava vektorů $\vec{u}^1, \ldots, \vec{u}^k$ se nazývá *ortogonální*, když

$$i \neq j \Longrightarrow (\vec{u}^i, \vec{u}^j) = 0.$$

Poznámka. Každá ortogonální soustava vektorů je lineárně nezávislá: Jsou-li $\vec{u}^1, \ldots, \vec{u}^k$ ortogonální, pak

$$\begin{aligned} c_1\vec{u}^1 + \ldots + c_k\vec{u}^k &= \vec{o} \quad |\cdot \vec{u}^i \\ c_1(\vec{u}^1, \vec{u}^i) + \ldots + c_k(\vec{u}^k, \vec{u}^i) &= (\vec{o}, \vec{u}^i) \\ c_i(\vec{u}^i, \vec{u}^i) &= (\vec{o}, \vec{u}^i) \\ c_i &= 0 \end{aligned}$$

pro $i = 1, \ldots, k$.

Věta 3. Ke každé symetrické matici A řádu n lze najít ortogonální soustava n vlastních vektorů. Tyto vektory tedy tvoří bazi v R^n.

8.1 Mocninná metoda

Základní varianta této metody najde aproximaci v absolutní hodnotě největšího vlastního čísla a příslušného vlastního vektoru dané matice A řádu n.

PŘEDPOKLADY:

1. Matice A je symetrická
2. Vlastní čísla matice A lze očíslovat $\lambda_1, \lambda_2, \ldots, \lambda_n$ tak, že $|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|$

Označíme $\vec{u}^i$ vlastní vektor příslušný vlastnímu číslu λ_i tak, že $\vec{u}^1, \ldots, \vec{u}^n$ tvoří ortogonální soustavu a $\|\vec{u}^i\| = 1$ pro $i = 1, \ldots, n$.

PRINCIP MOCNINNÉ METODY: Vektor $\vec{z}^0$ se zvolí a postupně se počítají vektory

$$\vec{z}^1 = A\vec{z}^0, \quad \vec{z}^2 = A\vec{z}^1, \ \ldots, \ \vec{z}^k = A\vec{z}^{k-1}, \ldots$$

Tedy pro $k = 1, 2, \ldots$ platí $\vec{z}^k = A^k\vec{z}^0$ a protože vektory $\vec{u}^1, \ldots, \vec{u}^n$ tvoří bazi v R^n, existují koeficienty $c_1, c_2, \ldots, c_n$:

$$\begin{aligned} \vec{z}^0 &= c_1\vec{u}^1 + c_2\vec{u}^2 + \ldots + c_n\vec{u}^n \quad \text{Pak} \\ \vec{z}^k = A^k\vec{z}^0 &= c_1A^k\vec{u}^1 + c_2A^k\vec{u}^2 + \ldots + c_nA^k\vec{u}^n \\ &= c_1\lambda_1^k\vec{u}^1 + c_2\lambda_2^k\vec{u}^2 + \ldots + c_n\lambda_n^k\vec{u}^n \\ &= \lambda_1^k\left(c_1\vec{u}^1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\vec{u}^2 + \ldots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\vec{u}^n\right) \end{aligned}$$

Tedy

$$\vec{z}^k \doteq \lambda_1^k c_1 \vec{u}^1 \tag{17}$$

pro dostatečně velká k, neboť

$$\left(\frac{\lambda_2}{\lambda_1}\right)^k \longrightarrow 0, \ldots, \left(\frac{\lambda_n}{\lambda_1}\right)^k \longrightarrow 0 \quad pro \quad k \longrightarrow \infty.$$

Tedy vektor $\vec{z}^k$ je, jako násobek vlastního vektoru $\vec{u}^1$, aproximací vlastního vektoru, příslušného vlastnímu číslu λ_1. Pak Rayleighův podíl

$$\sigma_k = \frac{\vec{z}^{k\,T} A \vec{z}^k}{\|\vec{z}^k\|_E^2} \doteq \lambda_1. \tag{18}$$

podle poznámky.

Z předchozí úvahy je patrné, že

(A) vektor $\vec{z}^0$ je třeba volit tak, aby koeficient c_0 byl co největší. V nejhorším případě, kdy $c_0 = 0$, platí

$$\vec{z}^0 \perp \vec{u}^1 \longrightarrow \vec{z}^k \perp \vec{u}^1$$

pro $k = 1, 2, \ldots$, takže $\vec{z}^k \not\longrightarrow \vec{u}^1$. Při praktickém výpočtu sice vlivem zaokrouhlovacích chyb zpravidla $\vec{z}^k \longrightarrow \vec{u}^1$, ale konvergence je velmi pomalá.

(B) Z (17) plyne, že

$$\begin{aligned} |\lambda_1| > 1 \quad &\Longrightarrow \quad \|\vec{z}^k\| \longrightarrow \infty \quad pro \quad k \longrightarrow \infty \\ |\lambda_1| < 1 \quad &\Longrightarrow \quad \|\vec{z}^k\| \longrightarrow 0 \quad pro \quad k \longrightarrow \infty \end{aligned}$$

Tyto implikace naznačují, že při tomto výpočtu pro velké hodnoty k vzniká nebezpečí přetečení v prvním případě a podtečení ve druhém případě. I když tyto extrémy nenastanou, vede práce s extrémně velkými případně extrémně malými hodnotami ke ztrátě přesnosti.

Důsledky: Z (A) plyne, že vektor $\vec{z}^0$ je třeba zvolit tak, aby jeho směr byl co nejblíže směru vektoru $\vec{u}^1$. Nebezpečí signalizované úvahou (B) bude odstraněno touto úpravou výpočtu: Každý nově vypočtený vektor $\vec{z}^k$ se normalizuje: Položí se

$$\vec{y}^k = c\vec{z}^k \quad tak,\ aby \quad \|\vec{y}^k\|_E = |c|\|\vec{z}^k\|_E = 1, \quad tj. \quad c = \frac{1}{\|z^k\|_E}.$$

KRITÉRIUM PRO UKONČENÍ: Zvolí se $\varepsilon > 0$ a výpočet se ukončí, jakmile

$$|\sigma_k - \sigma_{k-1}| \leq \varepsilon.$$

ALGORITMUS MOCNINNÉ METODY:

Nechť jsou dány symetrická matice A, vektor $\vec{z}^0$ a malé kladné číslo ε.

1. $\vec{y}^0 = \frac{1}{\|\vec{z}^0\|_E}\vec{z}^0$ (normalizace)

   $\vec{z}^1 = A\vec{y}^0$, $\sigma_0 = (\vec{y}^0, \vec{z}^1)$

2. Pro $k = 1, 2, \ldots$

   $\vec{y}^k = \frac{1}{\|\vec{z}^k\|_E}\vec{z}^k$ (normalizace)

   $\vec{z}^{k+1} = A\vec{y}^k$, $\sigma_k = (\vec{y}^k, \vec{z}^{k+1})$, dokud $|\sigma_k - \sigma_{k-1}| > \varepsilon$.

3. Je-li $|\sigma_k - \sigma_{k-1}| \leq \varepsilon$, pak $\vec{y} = \frac{1}{\|\vec{z}^{k+1}\|_E}\vec{z}^{k+1}$

   Výstup σ_k (aproximace λ_1) a $\vec{y}$ (aproximace $\vec{u}^1$)

Příklad 1. Aproximujte v absolutní hodnotě největší vlastní číslo matice

$$A = \begin{bmatrix} -4 & 1 \\ 1 & 2 \end{bmatrix}$$

a příslušný vlastní vektor. Položte $\vec{z}^0 = [-1, 0]^\top$ a $\varepsilon = 0.5 \cdot 10^{-3}$.

k	y_1^k	y_2^k	z_1^k	z_2^k	σ_k
0	-1	0	-1	0	-4
1	0,97014	-0,24254	4	-1	-4,11763
2	-0,99315	0,11684	4,12310	0,48506	-4,15016
	⋮				-4,15906
					-4,16138
					-4,16208
6	-0,98759	0,15703	-4,11060	0,65361	-4,16220
7	0,98682	-0,16182			

Tedy $\lambda_1 \doteq -4,16220$ a $\vec{u}^1 \doteq [0,98682, -0,16182]^\top$. Pro srovnání jsou přesné hodnoty

$$\lambda_1 = -4,16227766, \quad \vec{u}^1 = [0,987087, -0,16018224]^\top.$$

I z výsledku tohoto příkladu je patrná známá skutečnost, že hodnota vlastního čísla je zpravidla aproximována přesněji, než souřadnice příslušného vlastního vektoru.

Poznámka. Mocninné metody lze použít i pro aproximaci v absolutní hodnotě nejmenšího vlastního čísla a příslušného vlastního vektoru takto:

Podle Věty 1(d) má matice A^{-1} vlastní čísla

$$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}$$

a příslušné vlastní vektory $\vec{u}^1, \vec{u}^2, \ldots, \vec{u}^n$. Vzhledem k PŘEDPOKLADU 2 platí $|\frac{1}{\lambda_1}| \leq |\frac{1}{\lambda_2}| \leq \ldots \leq |\frac{1}{\lambda_n}|$. Jestliže $|\frac{1}{\lambda_{n-1}}| < |\frac{1}{\lambda_n}|$, pak matice A^{-1} splňuje PŘEDP. 1,2 a aplikace mocninné metody na matici A^{-1} poskytne aproximace $\sigma_k \doteq \frac{1}{\lambda_n}$ a $\vec{y} \doteq \vec{u}^n$.

Při aplikaci mocninné metody na matici A^{-1} se počítá

$$\vec{z}^{i+1} = A^{-1}\vec{y}^i \quad pro \quad i = 0, 1, \ldots \tag{19}$$

Aby nebylo nutno počítat matici A^{-1}, je místo (19) vhodnější řešit systém rovnic

$$A\vec{z}^{i+1} = \vec{y}^i \quad pro \quad i = 0, 1, \ldots$$

Zde se opakovaně řeší systémy rovnic s touž maticí A. Je výhodné poprvé najít LU-rozklad matice A a potom, pro $i = 1, 2, \ldots$, řešit úlohu $LU\vec{z}^{i+1} = \vec{y}^i$ pomocí dvou zpětných chodů.

9 Iterační metody řešení systémů lineárních algebraických rovnic

Budeme se opět zabývat řešením úlohy (8)

$$A\vec{x} = \vec{b}$$

za předpokladu, že matice A řádu n je pozitivně definitní, tj. že A je symetrická a $\vec{x}^\top A\vec{x} > 0$ pro všechna $\vec{x} \neq \vec{o}$. Označme $\hat{x}$ přesné řešení úlohy (8).

Věta 1. Všechna vlastní čísla pozitivně definitní matice A jsou kladná.

Důkaz. $A\vec{u} = \lambda\vec{u}$ a $\vec{u} \neq \vec{o} \Longrightarrow \lambda = \frac{\vec{u}^\top A\vec{u}}{\|\vec{u}\|_E^2} > 0$.

Úmluva. Vlastní čísla matice A označíme $\lambda_1, \lambda_2, \ldots, \lambda_n$ tak, aby $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ a odpovídající vlastní vektory $\vec{u}^1, \vec{u}^2, \ldots, \vec{u}^n$ zvolíme tak, aby tvořily ortogonální soustavu a aby $\|\vec{u}^i\|_E = 1$ pro $i = 1, 2, \ldots, n$.

Definice. Položíme

$$J(\vec{x}) = \frac{1}{2}\vec{x}^\top A\vec{x} - \vec{x}^\top\vec{b}.$$

Věta 2. (Hlavní věta) Je-li matice A pozitivně definitní, pak platí tato tvrzení (a), (b), (c):

(a) $J(\hat{x}) = -\frac{1}{2}\hat{x}^\top A\hat{x}$.

(b) $J(\vec{x}) = \frac{1}{2}(\vec{x} - \hat{x})^\top A(\vec{x} - \hat{x}) + J(\hat{x})$.

(c) $J(\hat{x}) < J(\vec{x})$ pro všechna $\vec{x} \neq \hat{x}$.

Důkaz (a): $J(\hat{x}) = \frac{1}{2}\hat{x}^\top A\hat{x} - \hat{x}^\top A\hat{x} = -\frac{1}{2}\hat{x}^\top A\hat{x}$.

Důkaz (b): $J(\vec{x}) = \frac{1}{2}\vec{x}^\top A\vec{x} - \vec{x}^\top A\hat{x}$

$$\begin{aligned} &= \frac{1}{2}\vec{x}^\top A(\vec{x}-\hat{x}) - \frac{1}{2}\vec{x}^\top A\hat{x} \\ &= \frac{1}{2}\vec{x}^\top A(\vec{x}-\hat{x}) - \frac{1}{2}(\vec{x}-\hat{x})^\top A\hat{x} - \frac{1}{2}\hat{x}^\top A\hat{x} \\ &= \frac{1}{2}(\vec{x}-\hat{x})^\top A(\vec{x}-\hat{x}) + J(\hat{x}). \end{aligned}$$

Důkaz (c): Tvrzení (c) plyne ihned z (b) a z pozitivní definitnosti matice A.

Definice. Nechť A je matice řádu n a $p \in \{1, 2, \ldots, n\}$. Symbolem $\vec{a}_p$ označíme p-tý řádkový vektor matice A.

Věta 3. Pro libovolnou čtvercovou matici A platí $grad\, J(\vec{x}) = A\vec{x} - \vec{b}$.

Důkaz. Lze snadno ověřit, že pro libovolný index $p \in \{1, 2, \ldots, n\}$ jsou všechny sčítance v $J(\vec{x})$, obsahující symbol x_p, právě

$$\frac{1}{2}x_p^2 a_{pp} + x_p \sum_{j \neq p} a_{pj} x_j - x_p b_p. \tag{20}$$

Tedy derivací $J(\vec{x})$ podle x_p vznikne

$$a_{pp}x_p + \sum_{j \neq p} a_{pj}x_j - b_p = \vec{a}_p\vec{x} - b_p. \tag{21}$$

9.1 Jacobiova metoda

pro danou k-tou aproximaci $\vec{x}^k$ a pro každý index $p \in \{1, 2, \ldots, n\}$ hledá novou hodnotu x_p^{k+1} jako tu hodnotu y, pro niž výraz

$$J_p^k = J(x_1^k, \ldots, x_{p-1}^k, y, x_{p+1}^k, \ldots, x_n^k)$$

nabývá svého minima. Z nutné podmínky pro minimum $\frac{\partial J}{\partial y} = 0$ a ze vztahů (20), (21) lze snadno odvodit tento vztah pro novou hodnotu x_p^{k+1}:

$$x_p^{k+1} = \frac{1}{a_{pp}} \left(b_p - \sum_{j \neq p} a_{pj} x_j^k \right) \quad pro \quad p = 1, 2, \ldots, n \tag{22}$$

Příklad 1. Systém rovnic $\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ 1 \\ -\frac{1}{3} \end{bmatrix}$ řešte Jacobiovou metodou.

$$\begin{aligned} x_1^{k+1} &= \frac{1}{2}\left(x_2^k + \frac{1}{3}\right) \\ x_2^{k+1} &= \frac{1}{2}\left(x_1^k + x_3^k + 1\right) \\ x_3^{k+1} &= \frac{1}{2}\left(x_2^k - \frac{1}{3}\right) \end{aligned}$$

Průběh výpočtu je zaznamenán v následující tabulce. Výpočet skončil po 26 krocích, jakmile $\vec{x}^{k+1} = \vec{x}^k$.

k	x_1^k	x_2^k	x_3^k
0	0	0	0
1	0,1667	0,5	-0,1667
2	0,4167	0,5	0,0833
$\vdots$	$\vdots$	$\vdots$	$\vdots$
25	0,6666	0,9999	0,3333
26	0,6666	0,9999	0,3333

Tato situace nemusí vždy nastat. Pro ukončení výpočtu se doporučuje toto kriterium: Zvolí se číslo $\varepsilon > 0$ a výpočet se ukončí, jakmile $\|\vec{x}^{k+1} - \vec{x}^k\| < \varepsilon$. Zde $\|\cdot\|$ je některá norma vektorů.

Poznámka. Jacobiova metoda konverguje pro každou počáteční aproximaci $\vec{x}^0$, je-li matice A pozitivně definitní.

Maticový zápis obecného předpisu (22) pro řešení úlohy (8) Jacobiovou metodou je

$$\vec{x}^{k+1} = C\vec{x}^k + \vec{d}, \quad \text{k}de$$

$$C = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \dots & -\frac{a_{2n}}{a_{22}} \\ \vdots & & & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & 0 \end{bmatrix} \quad \text{a} \quad \vec{d} = \begin{bmatrix} \frac{b_1}{a_{11}} \\ \frac{b_2}{a_{22}} \\ \vdots \\ \frac{b_n}{a_{nn}} \end{bmatrix}$$

Definice. Matice A řádu n se nazývá *silně diagonálně dominantní*, když

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad \text{pro} \quad i = 1, 2, \dots, n.$$

Věta 4. Je-li matice A silně diagonálně dominantní, pak Jacobiova metoda pro řešení $A\vec{x} = \vec{b}$ konverguje pro každou počáteční aproximaci $\vec{x}^0$.

Důkaz. $A\vec{x} = \vec{b} \Longleftrightarrow \vec{x} = C\vec{x} + \vec{d} \equiv F\vec{x}$ a platí

$$\|C\|_R = \max_{1\le i\le n} \sum_{j\neq i} \frac{|a_{ij}|}{|a_{ii}|} < 1.$$

Pak pro libovolné vektory $\vec{x}, \vec{y} \in R^n$ platí

$$\|F\vec{x} - F\vec{y}\|_R = \|C(\vec{x} - \vec{y})\|_R \le \|C\|_R \|\vec{x} - \vec{y}\|_R.$$

Tedy F je kontrakce v R^n s koeficientem $\|C\|_R < 1$. Tedy Jacobiova metoda konverguje podle Věty o kontrakci.

9.2 Gaussova-Seidelova metoda

Nechť jsou známy k-tá iterace $\vec{x}^k$ a z $(k+1)$-té iterace souřadnice $x_1^{k+1}, \ldots, x_{p-1}^{k+1}$. Pak hledáme p-tou komponentu x_p^{k+1} vektoru $\vec{x}^{k+1}$ jako tu hodnotu y, pro niž je výraz

$$\tilde{J}_p^k(y) = J(x_1^{k+1}, \ldots, x_{p-1}^{k+1}, y, x_{p+1}^k, \ldots, x_n^k)$$

minimální. Podobně jako v (22) lze užitím vztahů (20), (21) ukázat, že nutná podmínka $\frac{\partial \tilde{J}}{\partial y} = 0$ pro minimum $\tilde{J}$ poskytne hodnotu

$$x_p^{k+1} = \frac{1}{a_{pp}} \left(b_p - \sum_{j<p} a_{pj} x_j^{k+1} - \sum_{j>p} a_{pj} x_j^k \right) \tag{23}$$

Poznámka. Gaussova-Seidelova metoda konverguje pro všechny nulté aproximace $\vec{x}^0$, je-li matice A pozitivně definitní. V těchto případech konverguje rychleji, než metoda Jacobiova.

Příklad 2. Úlohu z př.1 řešte metodou Gaussovou-Seidelovou.

$$\begin{aligned} x_1^{k+1} &= \frac{1}{2}\left(x_2^k + \frac{1}{3}\right) \\ x_2^{k+1} &= \frac{1}{2}\left(x_1^{k+1} + x_3^k + 1\right) \\ x_3^{k+1} &= \frac{1}{2}\left(x_2^{k+1} - \frac{1}{3}\right) \end{aligned}$$

Průběh výpočtu je zaznamenán v následující tabulce. Výpočet skončil po 15

krocích, jakmile $\vec{x}^{k+1} = \vec{x}^k$.

k	x_1^k	x_2^k	x_3^k
0	0	0	0
1	0,1667	0,5833	0,1250
2	0,4583	0,7917	0,2290
⋮	⋮	⋮	⋮
14	0,6666	1,0000	0,3333
15	0,6666	1,0000	0,3333

9.3 Relaxační metoda

je definována relací

$$x_p^{k+1} = x_p^k + \frac{\omega}{a_{pp}} \left(b_p - \sum_{j<p} a_{pj} x_j^{k+1} - \sum_{j \geq p} a_{pj} x_j^k \right) \tag{24}$$

Poznámka. Pro konvergenci relaxační metody je nutné, aby $\omega \in (0, 2)$. Je-li $\omega > 1$, mluvíme o *superrelaxaci* a v případě $\omega < 1$ se metoda nazývá *subrelaxace*. Rychlost konvergence je velmi citlivá na volbu parametru ω.

Příklad 3. Úlohu z př.1 řešte relaxační metodou s parametrem $\omega = 1,2$.

$$\begin{aligned} x_1^{k+1} &= x_1^k + 1,2 \left(-x_1^k + \frac{1}{2} x_2^k + \frac{1}{6} \right) \\ x_2^{k+1} &= x_2^k + 1,2 \left(\frac{1}{2} x_1^{k+1} - x_2^k + \frac{1}{2} x_3^k + \frac{1}{2} \right) \\ x_3^{k+1} &= x_3^k + 1,2 \left(\frac{1}{2} x_2^{k+1} - x_3^k - \frac{1}{6} \right) \end{aligned}$$

Průběh výpočtu je zaznamenán v následující tabulce. Výpočet skončil po osmi krocích, jakmile $\vec{x}^{k+1} = \vec{x}^k$.

k	x_1^k	x_2^k	x_3^k
0	0	0	0
1	0,2	0,6	-0,2
2	0,52	0,672	0,2432
⋮	⋮	⋮	⋮
7	0,6667	1,0001	0,3334
8	0,6667	1,0001	0,3334

9.4 Metody největšího spádu, těžkého míče a konjugovaných gradientů

Další metody pro minimalizaci výrazu

$$J(\vec{x}) = \frac{1}{2}\vec{x}^\top A\vec{x} - \vec{x}^\top\vec{b}$$

pro pozitivně definitní matice A vychází z tohoto principu: Je dána k-tá aproximace $\vec{x}^k$ a směrový vektor $\vec{v}^k$. Hledá se koeficient α^k tak, aby pro

$$\vec{x}^{k+1} = \vec{x}^k + \alpha^k\vec{v}^k$$

platilo

$$J(\vec{x}^k + \alpha^k\vec{v}^k) \leq J(\vec{x}^k + \alpha\vec{v}^k) \quad pro\ vechna \quad \alpha \in R.$$

Hodnotu α^k lze snadno stanovit:

$$\begin{aligned}
\tilde{J}(\alpha) \equiv J(\vec{x}+\alpha\vec{v}) &= \frac{1}{2}(\vec{x}+\alpha\vec{v})^\top A(\vec{x}+\alpha\vec{v}) - (\vec{x}+\alpha\vec{v})^\top\vec{b} \\
&= \frac{1}{2}\vec{x}^\top A\vec{x} - \vec{x}^\top\vec{b} + \alpha[\vec{v}^\top A\vec{x} - \vec{v}^\top\vec{b}] + \frac{1}{2}\alpha^2\vec{v}^\top A\vec{v} \\
\frac{d\tilde{J}}{d\alpha} = \frac{dJ}{d\alpha}(\vec{x}+\alpha\vec{v}) &= \vec{v}^\top(A\vec{x} - \vec{b}) + \alpha\vec{v}^\top A\vec{v} = 0 \\
&\iff \quad \alpha = \frac{\vec{v}^T\cdot\vec{r}}{\vec{v}^\top A\vec{v}} \quad pro \quad \vec{r} = \vec{b} - A\vec{x}.
\end{aligned}$$

Položíme tedy

$$\alpha^k = \frac{\vec{v}^{k\,T}\cdot\vec{r}^k}{\vec{v}^{k\,T}A\vec{v}^k}\ , \quad kde \quad \vec{r}^k = \vec{b} - A\vec{x}^k \quad se\ nazv\ reziduum.$$

9.4.1 Metoda největšího spádu, gradientní metoda

Zvolí se $\vec{v}^k = -\,grad\,J(\vec{x}^k) = \vec{b} - A\vec{x}^k = \vec{r}^k$ podle Věty 3. Je dobře známo, že tento vektor je kolmý k vrstevnici a určuje směr největšího spádu hodnot výrazu $J(\vec{x})$. Postup výpočtu:

$$\begin{aligned}
&\vec{x}^0 \quad se\ zvol \quad \text{a}\ pro \quad k = 1, 2, \ldots \\
&\vec{x}^{k+1} = \vec{x}^k + \alpha^k\vec{r}^k\ , \quad kde \quad \alpha^k = \frac{\vec{r}^{k\,T}\cdot\vec{r}^k}{\vec{r}^{k\,T}A\vec{r}^k}
\end{aligned}$$

Poznámka. Počet kroků metody největšího spádu odpovídá číslu cond$A$.

Příklad 4. $\begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ 1 \\ -\frac{1}{3} \end{bmatrix}$

k	x_1^k	x_2^k	x_3^k	r_1^k	r_2^k	r_3^k	α^k
0	0	0	0	1/3	1	-1/3	0,5
1	0,1667	0,5	-0,1667	0,5	0	0,5	0,5
2	0,4167	0,5	0,0833	0	0,5	0	0,5
3	0,4167	0,75	0,0833	0,25	0	0,25	0,5
4	0,5417	0,75	0,2083	0	0,25	0	0,5
5	0,5417	0,875	0,2083				

9.4.2 Metoda těžkého míče

Pohyb míče po ploše v gravitačním poli není ve směru největšího spádu, ale závisí i na "starém směru":

$$\vec{v}^k = \vec{r}^k + \beta^{k-1}\vec{v}^{k-1},$$

kde $\beta^{k-1} \geq 0$ je vhodně zvolený parametr a $\vec{v}^{-1} = \vec{o}$.

9.4.3 Metoda konjugovaných gradientů

Definice. Nechť A je pozitivně definitní matice. Pro libovolné vektory $\vec{x}$, $\vec{y}$ položíme

$$\langle \vec{x}, \vec{y} \rangle_A = \vec{x}^\top A \vec{y}.$$

Věta 1. Předpis $\langle \cdot, \cdot \rangle_A$ je skalární součin v R^n.

Důkaz.

S1: $\langle \vec{x}, \vec{x} \rangle_A \geq 0$ a $\langle \vec{x}, \vec{x} \rangle_A = 0 \iff \vec{x} = \vec{o}$ plyne ihned z definice pozitivně definitní matice.

S2: $\langle \vec{x}, \vec{y} \rangle_A = \vec{x}^\top A \vec{y} = \vec{y} A^\top \vec{x} = \vec{y} A \vec{x} = \langle \vec{y}, \vec{x} \rangle_A$

S3: $\langle \vec{x}, a\vec{y} + g\vec{z} \rangle_A = a\vec{x}^\top A \vec{y} + b\vec{x}^\top A \vec{z} = a\langle \vec{x}, \vec{y} \rangle_A + b\langle \vec{x}, \vec{z} \rangle_A$.

Definice. Vektory $\vec{x}$, $\vec{y}$ se nazývají *konjugované*, když $\langle \vec{x}, \vec{y} \rangle_A = 0$.

Metoda konjugovaných gradientů je metoda těžkého míče, v níž je koeficient β^k zvolen tak, aby $\vec{v}^k$ a $\vec{v}^{k+1}$ byly konjugované vektory. Tedy algoritmus metody konjugovaných gradientů lze zapsat takto:

$\vec{x}^0$ se *zvol a pro* $k = 0, 1, \ldots$

$\vec{x}^{k+1} = \vec{x}^k + \alpha^k \vec{v}^k$, kde $\alpha^k = \dfrac{\vec{v}^{k\,T} \cdot \vec{r}^k}{\langle \vec{v}^k, \vec{v}^k \rangle_A}$. P*itom*

$$\vec{v}^0 = \vec{r}^0 = \vec{b} - A\vec{x}^0,$$
$$\vec{v}^k = \vec{r}^k + \beta^{k-1}\vec{v}^{k-1} \quad \text{a} \quad \beta^{k-1} = -\frac{\langle \vec{v}^{k-1}, \vec{r}^k \rangle_A}{\langle \vec{v}^{k-1}, \vec{v}^{k-1} \rangle_A} \quad pro \quad k = 1, 2, \ldots$$

Příklad 5. $\begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ 1 \\ -\frac{1}{3} \end{bmatrix}$

k	x_1^k	x_2^k	x_3^k	r_1^k	r_2^k	r_3^k
0	0	0	0	1/3	1	-1/3
1	0,1667	0,5	-0,1667	0,5	0	0,5
2	0,7052	0,8461	0,1409	-0,231	0,1539	0,231
3	0,6667	1	0,3333			

k	v_1^k	v_2^k	v_3^k	α^k	β^k
0	1/3	1	-1/3	0,5	0,4091
1	0,6364	0,4091	0,3636	0,84	0,3447
2	-0,0652	0,2605	0,3258	0,5907	

Poznámka. Je dokázáno, že metoda konjugovaných gradientů poskytne přesné řešení po n krocích (n je řád matice A). Prakticky jsou však aproximace dostatečně přesné po menším počtu iterací. Počet potřebných kroků odpovídá $\sqrt{condA}$.

10 Řešení systémů nelineárních rovnic

Z důvodu formální jednoduchosti se omezíme jen na systémy dvou nelineárních rovnic pro dvě neznámé. Budeme tedy pracovat v prostoru R^2 s některou z norem $\|\cdot\|_R$, $\|\cdot\|_S$, $\|\cdot\|_E$. Vektor $\vec{x} \in R^2$ budeme nazývat *bod* a značit $\vec{x} = (x, y)$ místo $\vec{x} = [x_1, x_2]^\top$. Libovolnou ohraničenou, otevřenou a souvislou podmnožinu v R^2 nazveme *oblastí* v R^2.

ÚLOHA. Nechť f, g jsou spojité funkce, definované na některé oblasti Ω_0 v R^2. Najděte bod $(x, y) \in \Omega_0$, který splňuje rovnice

$$\begin{aligned} f(x, y) &= 0 \\ g(x, y) &= 0 \end{aligned} \tag{25}$$

Příklad 1. Otázka existence a počtu řešení úlohy (25) je velmi složitá. Pro ilustraci této skutečnosti se přesvědčte, že systém rovnic

$$\begin{aligned} x^2 - y + a &= 0 \\ -x + y^2 + a &= 0 \end{aligned}$$

má jediné řešení pro $a = 0,25$, dvě řešení pro $a = 0$ a čtyři řešení pro $a = -1$.

10.1 Metoda prosté iterace

spočívá v převedení systému rovnic (25) na soustavu

$$\begin{aligned} x &= F(x,y) \\ y &= G(x,y) \end{aligned} \tag{26}$$

ekvivalentní s (25) na některé podoblasti Ω v Ω_0. Označíme-li

$$\vec{x} = (x,y) \quad \text{a} \quad \vec{F}(\vec{x}) = (F(x,y), G(x,y))\,,$$

vznikne vektorový zápis

$$\vec{x} = \vec{F}(\vec{x}),$$

formálně shodný se zápisem $x = F(x)$, používaným při řešení jedné rovnice pro jednu neznámou metodou prosté iterace. I v tomto případě tedy počáteční aproximaci $\vec{x}^0$ zvolíme a počítáme

$$\vec{x}^{k+1} = \vec{F}(\vec{x}^k).$$

Z obecných úvah z Kap.3 víme, že pokud posloupnost postupných aproximací $\{\vec{x}^k\}_0^\infty$ konverguje k bodu $\hat{x}$, je $\hat{x}$ řešením úlohy (26) a tedy i řešením úlohy (25).

Poznámka 1. (Kritérium pro ukončení) Použije se jedno z těchto kritérií:

a) zvolí se $\varepsilon > 0$ a výpočet se ukončí, jakmile $\|\vec{x}^{k+1} - \vec{x}^k\| < \varepsilon$

b) zvolí se a $\delta > 0$ a výpočet se ukončí, jakmile $\|\vec{F}(\vec{x}^k)\| < \delta$.

Poznámka 2. Funkce F a G je třeba volit tak, aby norma Jacobiovy matice

$$J(\vec{x}) = \begin{bmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{bmatrix}$$

byla co nejmenší. Lze snadno ověřit, že vektorová funkce $\vec{F} : R^2 \longrightarrow R^2$ je kontrakce na některé oblasti $\Omega_1 \subset \Omega$, když $\vec{F}$ zobrazuje Ω_1 do Ω_1 a některá norma splňuje podmínku $\max_{\vec{x}\in\Omega_1} \|J(\vec{x})\| < 1$. Podle Věty o kontrakci je pak konvergence tím rychlejší, čím menší toto maximum je.

Příklad 1. Grafickou metodou určete počet kořenů systému rovnic

$$\begin{aligned} f(x,y) \equiv xy - y - 1 &= 0 \\ g(x,y) \equiv x^2 - y^2 - 1 &= 0 \end{aligned}$$

a hrubé odhady jejich hodnot. Metodou prosté iterace určete jeden z kořenů s chybou menší, než 0,0005.

Protože

$$f(x,y)=0 \Longrightarrow y=\frac{1}{x-1} \quad \text{a} \quad g(x,y)=0 \Longrightarrow x=\pm\sqrt{y^2+1}, \qquad (27)$$

lze schematickým znázorněním grafů těchto funkcí zjistit, že úloha má dvě řešení, jejichž přibližné hodnoty jsou

$$\vec{x}^1=(1,7,1,3) \quad \text{a} \quad \vec{x}^2=(-1,1,-0,2).$$

Tedy Jacobiova matice vektorové funkce $\vec{F}=(\sqrt{y^2+1},\frac{1}{x-1})$ je

$$J(\vec{x})=\begin{bmatrix} 0 & \frac{y}{\sqrt{y^2+1}} \\ -\frac{1}{(x-1)^2} & 0 \end{bmatrix}$$

a její řádková i sloupcová norma má hodnotu $\|J(\vec{x})\|=\max\{\frac{|y|}{\sqrt{y^2+1}},\frac{1}{(x-1)^2}\}$. Protože $\|J(\vec{x}^1)\| \doteq 2,041$ a $\|J(\vec{x}^2)\| \doteq 0,2268$, použijeme předpisu (27) pro zpřesnění kořene $\vec{x}^2$. Předpis má tedy tvar

$$\begin{aligned} \vec{x}^0 &= (x^0,y^0)=(-1,1,-0,2) \quad \text{a} \\ \vec{x}^{k+1} &= (x^{k+1},y^{k+1})=(-\sqrt{(y^k)^2+1},\frac{1}{x^k-1}) \quad pro \quad k=0,1,\ldots \end{aligned}$$

Výpočet se ukončí, jakmile $\|\vec{x}^{k+1}-\vec{x}^k\|_R<0,0005$. Průběh výpočtu je zaznamenán v této tabulce:

i	x^i	y^i
0	-1,1	-0,2
1	-1,0198	-0,4762
2	-1,1076	-0,4951
3	-1,1159	-0,4745
4	-1,1069	-0,4726
5	-1,1061	-0,4746
6	-1,1069	-0,4748
7	-1,1070	-0,4746

10.2 Newtonova metoda (metoda linearizace)

Předpokládejme, že funkce f, g mají spojité druhé parciální derivace a že známe aproximaci $\vec{x}^k=(x^k,y^k)$ *blízko* přesného řešení $\hat{x}=(\hat{x},\hat{y})$ úlohy (25). Aproximujeme-li nulové hodnoty $f(\hat{x})$, $g(\hat{x})$ Taylorovým polynomem prvního stupně v okolí bodu $\vec{x}^k$, vznikne

$$\begin{aligned} f(\vec{x}^k)+\frac{\partial f}{\partial x}(\vec{x}^k)(\hat{x}-x^k)+\frac{\partial f}{\partial y}(\vec{x}^k)(\hat{y}-y^k) &\doteq 0 \\ g(\vec{x}^k)+\frac{\partial g}{\partial x}(\vec{x}^k)(\hat{x}-x^k)+\frac{\partial g}{\partial y}(\vec{x}^k)(\hat{y}-y^k) &\doteq 0 \end{aligned} \qquad (28)$$

Rozdíl mezi levou a pravou stranou rovnic v (28) je úměrný součinům

$$(\hat{x}-x^k)^2,(\hat{x}-x^k)(\hat{y}-y^k),(\hat{y}-y^k)^2,$$

což jsou za předpokladu, že aproximace $\vec{x}^k = (x^k, y^k)$ leží blízko přesného řešení $\hat{x} = (\hat{x}, \hat{y})$, velmi malé hodnoty. Nahradíme-li v (28) bod $(\hat{x}, \hat{y})$ bodem (x^{k+1}, y^{k+1}) a požadujeme-li přesné splnění rovností (28), vznikne systém rovnic

$$J(\vec{x}^k) \begin{bmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{bmatrix} = - \begin{bmatrix} f(\vec{x}^k) \\ g(\vec{x}^k) \end{bmatrix}, \tag{29}$$

což je jeden krok řešení úlohy (25) *Newtonovou metodou.*

Poznámka. V Newtonově metodě se aproximace $\vec{x}^k$ blíží k přesnému řešení $\hat{x}$ rychle, ale jen tehdy, když je nultá aproximace $\vec{x}^0$ dostatečně blízko k přesnému řešení $\hat{x}$.

Příklad 2. Aproximaci kořene úlohy z příkladu 1 zpřesněte co nejvíce Newtonovou metodou.

Položíme tedy $\vec{x}^0$=(-1,1070,-0,4746). Pak pro

$$J(\vec{x}) = \begin{bmatrix} y & x-1 \\ 2x & -2y \end{bmatrix}$$

platí

$$J(\vec{x}^0) = \begin{bmatrix} -0,4746 & -2,1070 \\ -2,2140 & 0,9492 \end{bmatrix} \text{ a } \quad - \begin{bmatrix} f(\vec{x}^0) \\ g(\vec{x}^0) \end{bmatrix} = \begin{bmatrix} 0,0000178 \\ -0,00020384 \end{bmatrix}.$$

řešením tohoto případu systému rovnice (28) získáme

$$\vec{x}^1 - \vec{x}^0 = \begin{bmatrix} 0,000080658 \\ -0,000266161 \end{bmatrix},$$

takže $\vec{x}^1 = \begin{bmatrix} -1,106919342 \\ -0,474626616 \end{bmatrix}$. Stejným postupem byly vypočteny i aproximace $\vec{x}^2$ a $\vec{x}^3$ z níže uvedené tabulky.

i	x^i	y^i
0	-1,1070	-0,4746
1	-1,106919342	-0,474626616
2	-1,106919340	-0,474626618
3	-1,106919340	-0,474626618

11 Funkční prostory

Definice. Pro libovolné funkce f, g se stejným definičním oborem D a pro $\alpha, \beta \in R$ se funkce

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x)$$

nazývá *lineární kombinace* funkcí f, g s koeficienty α, β.

Definice. *Funkční prostor* je každá neprázdná množina funkcí $\mathcal{F}$ se stejným definičním oborem s vlastností

$$f, g \in \mathcal{F} \Longrightarrow \alpha f + \beta g \in \mathcal{F} \quad pro\ vechna \quad \alpha, \beta \in R.$$

11.1 Příklady funkčních prostorů

Uvedeme několik příkladů funkčních prostorů i jejich konstrukcí.

Příklad 1. Množina $C\langle a, b\rangle$ všech funkcí spojitých na intervalu $\langle a, b\rangle$ je funkční prostor.

Příklad 2. $C^{(k)}\langle a, b\rangle = \{f; f, f', \ldots, f^{(k)} \in C\langle a, b\rangle\}$ je funkční prostor pro $k = 1, 2, \ldots$.

Příklad 3. $\{f \in C\langle a, b\rangle; f(x) > 0 \quad pro\ vechna \quad x \in \langle a, b\rangle\}$ není funkční prostor.

Zřejmě platí

$$C\langle a, b\rangle \supset C^{(1)}\langle a, b\rangle \supset C^{(2)}\langle a, b\rangle \supset \ldots \supset C^{(\infty)}\langle a, b\rangle,$$

kde

$$C^{(\infty)}\langle a, b\rangle = \bigcap_{k=1}^{\infty} C^{(k)}\langle a, b\rangle$$

a každé dva z těchto prostorů jsou vzájemně různé.

Definice. Je-li $\mathcal{F}$ prostor funkcí, definovaných na $\langle a, b\rangle$ (na oblasti $\bar{\Omega}$), položíme

$$\mathcal{F}_0 = \{f \in \mathcal{F}; f(a) = 0 = f(b)\} \qquad (\mathcal{F}_0 = \{f \in \mathcal{F}; f(x, y) = 0 \quad na\ hranici \quad \Omega\})$$

Ověřte, že $\mathcal{F}_0$ je vždy funkční prostor.

Definice. Funkce $f_1, \ldots, f_k$ z funkčního prostoru $\mathcal{F}$ jsou *lineárně nezávislé*, jestliže

$$c_1 f_1 + \ldots + c_k f_k = o$$

jen tehdy, když $c_1 = c_2 = \ldots = c_k = 0$. Každá maximální lineárně nezávislá množina funkcí v $\mathcal{F}$ se nazývá *báze* prostoru $\mathcal{F}$.

Věta 1. Funkce $\varphi_0(x) = 1$, $\varphi_1(x) = x$, …, $\varphi_n(x) = x^{n-1}$, definované na intervalu $\langle a, b\rangle$ jsou lineárně nezávislé pro každé $n > 0$.

Definice. Označíme $\psi_i(x) = (x - a)(x - b)\varphi_i(x)$ pro $i = 0, 1, \ldots$ a pro všechna $x \in \langle a, b\rangle$.

Věta 2. Funkce $\psi_0, \psi_1, \ldots, \psi_n$ jsou lineárně nezávislé pro všechna $n > 0$.

Poznámka. Lze snadno ověřit, že libovolná soustava polynomů vzájemně různých stupňů je lineárně nezávislá.

Poznámka. Funkce $\varphi_0, \varphi_1, \ldots, \varphi_n$ leží ve všech prostorech $\mathcal{F}$ z příkladů 1-3 a funkce $\psi_0, \psi_1, \ldots, \psi_n$ leží ve všech prostorech $\mathcal{F}_0$ z příkladů 1-3.

Tedy ani prostory $\mathcal{F}$ z Příkladů 1–3, ani prostory $\mathcal{F}_\prime$ z Příkladů 1–3 nemají konečnou dimenzi. Říkáme, že jejich dimenze je nekonečná.

Definice. Pro libovolné funkce $f_1, f_2, \ldots, f_n$ s týmž definičním oborem označíme $\mathcal{L}(f_1, \ldots, f_n)$ množinu všech lineárních kombinací funkcí $f_1, f_2, \ldots, f_n$.

Věta 3. $\mathcal{L}(f_1, \ldots, f_n)$ je funkční prostor. Jsou-li $f_1, f_2, \ldots, f_n$ navíc lineárně nezávislé, pak $f_1, f_2, \ldots, f_n$ tvoří bazi v prostoru $\mathcal{L}(f_1, \ldots, f_n)$.

Definice. Položíme

$$\begin{aligned} \mathcal{P}^0 &= \mathcal{L}(1), \\ \mathcal{P}^1 &= \mathcal{L}(1, x), \\ &\vdots \\ \mathcal{P}^n &= \mathcal{L}(1, x, \ldots, x^n) \quad \text{a} \\ \mathcal{P} &= \bigcup_{n=0}^{\infty} \mathcal{P}^n \end{aligned}$$

Definice. Vzájemně různá reálná čísla se nazývají *uzly*. Uzly x_0, x_1, …, x_n se nazývají *ekvidistantní*, existuje-li kladné číslo h *krok* tak, že $x_i = x_0 + ih$ pro $i = 1, 2, \ldots, n$.

Příklad 5. (Prostor lineárních splajnů) Buďte $a = x_0 < x_1 < \ldots < x_n = b$ ekvidistantní uzly s krokem h. Označíme $\mathcal{S}^1(a, b, h)$ množinu všech funkcí φ splňujících tyto podmínky (a), (b):

(a) $\varphi \in C\langle a, b\rangle$,

(b) φ je lineární na $\langle x_{i-1}, x_i\rangle$ pro $i = 1, \ldots, n$.

Zřejmě platí

(c) Libovolná funkce $f \in \mathcal{S}^1(a, b, h)$ je jednoznačně určena hodnotami $f_i = f(x_i)$ pro $i = 0, \ldots, n$.

Tedy pro $i = 0, \ldots, n$ předpis

$$w_i(x_i) = 1 \quad \text{a} \quad w_i(x_j) = 0 \quad \textit{pro vechna} \quad j \neq i$$

určuje funkci w_i jednoznačně.

Dokažte, že platí

1. Funkce $f \equiv c_0 w_0 + c_1 w_1 + \ldots + c_n w_n$ má v uzlech hodnoty $f(x_i) = c_i$ pro $i = 0, 1, \ldots, n$

2. $c_0 w_0 + c_1 w_1 + \ldots + c_n w_n = o \Longrightarrow c_0 = c_1 = \ldots = c_n = 0$

3. $f \in \mathcal{S}^1(a, b, h) \Longrightarrow f = f(x_0)w_0 + f(x_1)w_1 + \ldots + f(x_n)w_n$ a tedy $\mathcal{S}^1(a, b, h) = \mathcal{L}(w_0, w_1, \ldots, w_n)$,

takže funkce $w_0, w_1, \ldots, w_n$ tvoří bazi v prostoru $\mathcal{S}^1(a, b, h)$.

11.2 Normy a skalární součin funkcí

$$f \in C\langle a, b\rangle \Longrightarrow |f| \in C\langle a, b\rangle \Longrightarrow \quad existuje \quad |f|_C = \max_{a \le x \le b} |f(x)|$$

Přiřazení $f \longmapsto |f|_C$ z $C\langle a, b\rangle$ do R má vlastnosti

N1 $|f|_C \ge 0$ a $|f|_C = 0 \Longleftrightarrow f = o$

N2 $|\alpha f|_C = |\alpha|\,|f|_C$

N3 $|f + g|_C \le |f|_C + |g|_C$,

takže $|\cdot|_C$ je *norma* na funkčním prostoru $C\langle a, b\rangle$. Nazývá se *čebyševovská norma*.

$$f \in C\langle a, b\rangle \Longrightarrow \quad existuje\ konen \quad \int_a^b f^2(x)\,dx.$$

Pak

$$\|f\| = \sqrt{\int_a^b f^2(x)\,dx}$$

má vlastnosti N1-N3, takže i $\|\cdot\|$ je norma na prostoru $C\langle a, b\rangle$. Nazývá se *euklidovská norma*.

Definice. Pro $f, g \in C\langle a, b\rangle$ položme

$$(f, g) = \int_a^b f(x)\,g(x)\,dx.$$

Věta 1. Přiřazení $(\cdot, \cdot)$ má vlastnosti S1–S4 z odstavce 6, takže je to skalární součin na $C\langle a, b\rangle$. Navíc zřejmě platí

$$\|f\|^2 = (f, f) \quad pro\ vechna \quad f \in C\langle a, b\rangle.$$

Předpisy $d_C(f, g) = |f - g|_C$ a $d_E(f, g) = \|f - g\|$ definují metriky na $C\langle a, b\rangle$.

Věta 2. Metrický prostor $(C\langle a, b\rangle, d_C)$ je úplný, ale metrický prostor $(C\langle a, b\rangle, d_E)$ není úplný.

Skutečnost, že prostor $(C\langle a,b\rangle, d_E)$ není úplný, lze snadno ukázat ověřením, že posloupnost $(f_n)_1^\infty$, kde

$$f_n(x) = \begin{cases} -1 \quad pro \quad x \in \langle -1, -\frac{1}{n}\rangle \\ nx \quad pro \quad x \in \langle -\frac{1}{n}, \frac{1}{n}\rangle \\ 1 \quad pro \quad x \in \langle \frac{1}{n}, 1\rangle \end{cases},$$

je cauchyovská v $(C\langle a,b\rangle, d_E)$, ale není v tomto prostoru konvergentní.

Definice. Symbolem $L_2(a,b)$ označíme množinu všech funkcí f, pro něž Lebesgueův integrál

$$\int_a^b f^2(x)\,dx$$

existuje a má konečnou hodnotu.

Poznámka. Pojem Lebesgueova integrálu se prakticky neliší od známého pojmu Riemannova integrálu. Existuje-li Riemannův integrál, pak existuje i Legesgueův integrál a mají stejnou hodnotu.

Definujeme-li v množině funkcí $L_2(a,b)$ rovnost předpisem

$$f = g \quad kdy \quad \int_a^b (f(x) - g(x))^2\,dx = 0,$$

pak platí

Věta 3. $(L_2(a,b), d_E)$ je nejmenší úplný metrický prostor s vlastností $L_2(a,b) \supseteq C\langle a,b\rangle$, $\|\cdot\|$ je normou v $L_2(a,b)$ a $(\cdot,\cdot)$ je skalární součin v $L_2(a,b)$.

12 Interpolace a aproximace funkce

ÚLOHA LAGRANGEOVY INTERPOLACE: Jsou dány uzly $x_0, x_1, \ldots, x_n$ a hodnoty $f_i = f(x_i)$ pro $i = 0, 1, \ldots, n$. Najděte "jednoduchou" funkci φ, splňující

$$\varphi(x_i) = f_i \quad pro \quad i = 0, 1, \ldots, n. \tag{30}$$

Funkce φ se nazývá *interpolant* funkce f v uzlech $x_0, x_1, \ldots, x_n$.

12.1 Interpolační polynomy

Věta 1. Nechť je dáno $n+1$ uzlů $x_0, x_1, \ldots, x_n$ a hodnoty $f_i = f(x_i)$ funkce f pro $i = 0, 1, \ldots, n$. Pak v $\mathcal{P}^n$ existuje jediný interpolant $P(x)$ funkce f v uzlech $x_0, x_1, \ldots, x_n$.

Důkaz. Každý polynom $P(x) \in \mathcal{P}^n$ lze zapsat ve tvaru $P(x) = a_0 + a_1 x + \ldots + a_n x^n$. Podmínky (30):

$$P(x_i) = f_i \quad pro \quad i = 0, 1, \ldots, n$$

lze zapsat ve tvaru

$$\begin{bmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ \vdots & & & \vdots \\ 1 & x_n & \dots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{bmatrix}. \tag{31}$$

Determinant matice V tohoto systému rovnic se nazývá Vandermondův a má hodnotu

$$\det V = \prod_{0 \le j < i \le n} (x_i - x_j).$$

Tedy $\det V \neq 0$, neboť $x_i \neq x_j$ pro všechna $i \neq j$ a to znamená, že úloha má jediné řešení.

Definice. Polynom $P(x)$ z Věty 1 se nazývá *interpolační polynom* funkce f v uzlech $x_0, x_1, \dots, x_n$.

KONSTRUKCE:

I. Interpolant v základním tvaru: V $\mathcal{P}^n$ se zvolí baze $\varphi_0 = 1$, $\varphi_1 = x$, …, $\varphi_n = x^{n-1}$ a koeficienty $a_0, a_1, \dots, a_n$ vzniknou řešením systému rovnic (31).

II. Interpolant v Newtonově tvaru: V $\mathcal{P}^n$ se zvolí baze $\varphi_0 = 1$, $\varphi_1 = (x - x_0)$, $\varphi_2 = (x - x_0)(x - x_1)$, …, $\varphi_n = (x - x_0)(x - x_1)\dots(x - x_{n-1})$. Polynom

$$P(x) = a_0 + a_1(x - x_0) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1})$$

splňuje podmínky (30) právě když

$$\begin{aligned} a_0 &= f_0 \\ a_0 + (x_1 - x_0)a_1 &= f_1 \\ \vdots &= \vdots \\ a_0 + (x_n - x_0)a_1 + \dots + (x_n - x_0)\dots(x_n - x_{n-1})a_n &= f_n \end{aligned} \tag{32}$$

Protože matice systému rovic (32) je dolní trojúhelníková, je jeho řešení podstatně efektivnější, než v případě (31). Navíc lze toto řešení popsat užitím rekurze takto: Zřejmě

$$a_0 = f_0 \quad \text{a} \quad a_1 = \frac{f_1 - f_0}{x_1 - x_0}.$$

Obecněji lze ukázat, že

$$a_i = f[x_0, x_1, \dots, x_i] \quad \text{pro} \quad i = 1, 2, \dots, n,$$

kde

$$\begin{aligned} f[x_i, x_{i+1}] &= \frac{f_{i+1} - f_i}{x_{i+1} - x_i} \quad pro \quad i = 0, \ldots, n-1 \\ f[x_i, x_{i+1}, x_{i+2}] &= \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i} \quad pro \quad i = 0, \ldots, n-2 \\ &\vdots \\ f[x_0, \ldots, x_n] &= \frac{f[x_1, \ldots, x_n] - f[x_0, \ldots, x_{n-1}]}{x_n - x_0}. \end{aligned}$$

Tyto výrazy se postupně nazývají *poměrné diference* 1.,2.,. . ., n-tého řádu. Tedy interpolant v Newtonově tvaru je polynom

$$\begin{aligned} P(x) &= f_0 + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \ldots \\ &+ f[x_0, x_1, \ldots, x_n](x - x_0)(x - x_1)\ldots(x - x_{n-1}) \end{aligned} \tag{33}$$

Příklad 1. Určete Newtonův interpolační polynom funkce f v uzlech z tabulky

i	x_i	f_i	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_0, x_1, x_2, x_3]$
0	-1	3	-1	0	1
1	0	2	-1	4	
2	1	1	11		
3	3	23			

Tedy $P(x) = 3 - (x + 1) + (x + 1)x(x - 1)$.

Poznámka. Konstrukce interpolačního polynomu v Newtonově tvaru je efektivní (v podstatě spočívá v řešení systému rovnic s trojúhelníkovou maticí oproti konstrukci I, kde je řešen systém stejného počtu rovnic s plnou maticí) a zároveň lze se získaným tvarem efektivně pracovat podobně jako s polynomem v základním tvaru. Příklad 2 ukazuje, že například algoritmus výpočtu hodnoty polynomu v Newtonově tvaru je v podstatě stejně efektivní (provádí stejný počet operací násobení) jako optimálně rychlý algoritmus Hornerova schematu pro polynomy v základním tvaru.

Příklad 2. Výpočet hodnoty polynomu ve tvaru (33) zobecněným Hornerovým schematem.

$p := f_n;$
$for \quad i := n - 1 \quad downto \quad 0 \quad do$
$\quad p := p * (x - x_i) + f_i;$

Věta 2. Nechť $f \in C^{(n+1)}\langle a,b\rangle$ a $P \in \mathcal{P}^n$ je interpolant funkce f v uzlech $a = x_0 < x_1 < \ldots < x_n = b$. Pak ke každému $x \in \langle a,b\rangle$ existuje $\xi \in (a,b)$ tak, že

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x),$$

kde $\omega(x) = (x - x_0)(x - x_1)\ldots(x - x_n)$.

Poznámka. Polynom $\omega(x)$ má pro x blízko a nebo b podstatně větší hodnoty, než blízko středu intervalu $\langle a,b\rangle$. V případě, že délka intervalu $\langle a,b\rangle$ je větší nebo uzlů interpolace je větší počet a jsou rozmístěny rovnoměrně, je chyba aproximace poblíž hranice intervalu $\langle a,b\rangle$ extrémně velká. Tento nedostatek (toto omezení použitelnosti) interpolačního polynomu se nazývá *Rungeho jev.*

12.2 Interpolační kubické splajny

Pro intervaly větší délky je použití interpolačního polynomu

nízkého stupně příliš nepřesné a

vyšších stupňů nevhodné vzhledem k Rungeho jevu.

Proto se $\langle a,b\rangle$ rozděluje na několik (krátkých) podintervalů a na každém se konstruuje obecně jiný interpolační polynom nízkého stupně. Výsledný interpolant je "po částech" polynom. Jde o to, aby takto vytvořený interpolant byl co nejhladší, tj. aby měl co nejvíce spojitých derivací.

Poznámka. Interpolační lineární splajn.

Definice. Nechť $a = x_0 < x_1 < \ldots < x_n = b$. *Kubický splajn* s uzly $x_0, \ldots, x_n$ je každá funkce $g(x)$ s těmito vlastnostmi (a), (b):

(a) Na $\langle x_{i-1}, x_i\rangle$ je $g(x) = g_i(x)$ polynom stupně nejvýše třetího a

(b) $g(x) \in C^{(2)}\langle a,b\rangle$, tj. g, g', g'' jsou spojité na $\langle a,b\rangle$.

Poznámka. Podmínka (b) je zřejmě ekvivalentní s podmínkou

$$\begin{aligned} g_i(x_i) &= g_{i+1}(x_i) \\ g_i'(x_i) &= g_{i+1}'(x_i) \qquad \text{pro } i = 1, \ldots, n-1 \\ g_i''(x_i) &= g_{i+1}''(x_i) \end{aligned}$$

ÚLOHA. Nechť $a = x_0 < x_1 < \ldots < x_n = b$ a $f_i = f(x_i)$ pro $i = 0, 1, \ldots, n$. Hledáme kubický splajn $g(x)$ s uzly $x_0, \ldots, x_n$ tak, aby

$$g(x_i) = f_i \quad pro \quad i = 0, 1, \ldots, n. \tag{34}$$

Pak se g nazývá *interpolační kubický splajn* funkce f v uzlech $x_0, \ldots, x_n$.

Poznámka. Ukážeme, že jestliže se navíc požaduje

$$g''(a) = 0 = g''(b), \tag{35}$$

pak má úloha jediné řešení.

Definice. Kubické splajny, splňující podmínku (35), se nazývají *přirozené*.

ODVOZENÍ KONSTRUKCE INTERPOLAČNÍHO KUBICKÉHO SPLAJNU: Podle definice je pro každý kubický splajn g druhá derivace g'' spojitá na $\langle a, b\rangle$ a na $\langle x_{k-1}, x_k\rangle$ lineární pro $k = 1, \ldots, n$. Tedy existují koeficienty $m_0 = g_1''(x_0)$, …, $m_n = g_n''(x_n)$ tak, že platí

$$g_k''(x) = m_{k-1}\frac{x_k - x}{h_k} + m_k\frac{x - x_{k-1}}{h_k}, \tag{36}$$

kde $h_k = x_k - x_{k-1}$ pro $k = 1, \ldots, n$.

Funkci (36) integrujme dvakrát vzhledem k proměnné x:

$$g_k(x) = m_{k-1}\frac{(x_k - x)^3}{6h_k} + m_k\frac{(x - x_{k-1})^3}{6h_k} + A_k\frac{x_k - x}{h_k} + B_k\frac{x - x_{k-1}}{h_k}.$$

Hodnoty integračních konstant A_k a B_k určíme tak, aby byly splněny podmínky interpolace (34), čímž bude zároveň zajištěna spojitost splajnu g:

$$\begin{aligned} g_k(x_k) = m_k\frac{h_k^2}{6} + B_k = f_k \quad &\Longrightarrow \quad B_k = f_k - m_k\frac{h_k^2}{6} \\ g_k(x_{k-1}) = m_{k-1}\frac{h_k^2}{6} + A_k = f_{k-1} \quad &\Longrightarrow \quad A_k = f_{k-1} - m_{k-1}\frac{h_k^2}{6} \end{aligned}$$

a tedy

$$\begin{aligned} g_k(x) &= m_{k-1}\frac{(x_k - x)^3}{6h_k} + m_k\frac{(x - x_{k-1})^3}{6h_k} \\ &+ \left(f_{k-1} - m_{k-1}\frac{h_k^2}{6}\right)\frac{x_k - x}{h_k} + \left(f_k - m_k\frac{h_k^2}{6}\right)\frac{x - x_{k-1}}{h_k}. \end{aligned} \tag{37}$$

Zbývá zvolit hodnoty koeficientů $m_0, \ldots, m_n$ tak, aby byla spojitá i první derivace g':

$$\begin{aligned} g_k'(x) &= -m_{k-1}\frac{(x_k - x)^2}{2h_k} + m_k\frac{(x - x_{k-1})^2}{2h_k} + \frac{f_k - f_{k-1}}{h_k} \\ &- \frac{m_k - m_{k-1}}{6}h_k \\ g_k'(x_k) &= -m_{k-1}\frac{h_k}{6} + m_k\frac{h_k}{3} + \frac{f_k - f_{k-1}}{h_k} \\ g_{k+1}'(x_k) &= -m_k\frac{h_{k+1}}{3} - m_{k+1}\frac{h_{k+1}}{6} + \frac{f_{k+1} - f_k}{h_{k+1}} \end{aligned}$$

Odtud a z podmínky $g_k'(x_k) = g_{k+1}'(x_k)$ plyne rovnice

$$m_{k-1}\frac{h_k}{6} + m_k\frac{h_k + h_{k+1}}{3} + m_{k+1}\frac{h_{k+1}}{6} = \frac{f_{k+1} - f_k}{h_{k+1}} - \frac{f_k - f_{k-1}}{h_k} \tag{38}$$

pro $k = 1, \ldots, n-1$ a $m_0 = 0 = m_n$ plyne z (35).

Označíme-li

$$\vec{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix}, \quad \vec{f} = \begin{bmatrix} f_0 \\ \vdots \\ f_n \end{bmatrix}, \quad A = \begin{bmatrix} \frac{h_1+h_2}{3} & \frac{h_2}{6} & & & \\ \frac{h_2}{6} & \frac{h_2+h_3}{3} & \frac{h_3}{6} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{h_{n-1}}{6} & \frac{h_{n-1}+h_n}{3} \end{bmatrix}$$

$$\text{a} \quad H = \begin{bmatrix} \frac{1}{h_1} & -\frac{1}{h_1} - \frac{1}{h_2} & \frac{1}{h_2} & & \\ & \frac{1}{h_2} & -\frac{1}{h_2} - \frac{1}{h_3} & \frac{1}{h_3} & \\ & & \ddots & \ddots & \ddots \\ & & \frac{1}{h_{n-1}} & -\frac{1}{h_{n-1}} - \frac{1}{h_n} & \frac{1}{h_n} \end{bmatrix},$$

pak je maticový zápis systému rovnic (38):

$$A\vec{m} = H\vec{f} \tag{39}$$

KONSTRUKCE INTERPOLAČNÍHO KUBICKÉHO SPLAJNU:

1. Výpočet $h_1, \ldots, h_n$.
2. Sestavení matic A a H.
3. Řešení systému rovnic (38).
4. Dosazení do (37).

Níže uvedená věta je přesným matematickým vyjádřením skutečnosti, že kubický splajn je "nejhladším interpolantem".

Věta 3. Přirozený kubický splajn $g(x)$ je jediný interpolant funkce f v uzlech $x_0, \ldots, x_n$, který leží v $C^{(2)}\langle a, b\rangle$ takový, že

$$\Phi(g) \leq \Phi(\varphi)$$

pro všechny interpolanty $\varphi \in C^{(2)}\langle a, b\rangle$ funkce f v uzlech $x_0, \ldots, x_n$. Zde $\Phi(\varphi) = \int_a^b \varphi''^2(x)\,dx$.

12.3 Hermiteovy interpolační polynomy

ÚLOHA HERMITEOVY INTERPOLACE: Je dáno $n+1$ uzlů $x_0, \ldots, x_n$ a hodnoty $f_j = f(x_j)$, $f_j' = f'(x_j)$ funkce f a její derivace f' pro $j = 0, 1, \ldots, n$. Je třeba najít "co nejjednodušší funkci" φ tak, aby

$$\varphi(x_j) = f_j \quad \text{a} \quad \varphi'(x_j) = f_j' \quad pro \quad j = 0, 1, \ldots, n \tag{40}$$

Řekneme, že funkce φ je *Hermiteův interpolant* funkce f v uzlech $x_0, x_1, \ldots, x_n$.

(A) φ je polynom co nejnižšího stupně.

Věta 4. Nechť je dáno $n+1$ uzlů $x_0, x_1, \ldots, x_n$ a hodnoty $f_j = f(x_j)$, $f'_j = f'(x_j)$ pro $j = 0, 1, \ldots, n$. Pak v $\mathcal{P}^{2n+1}$ existuje právě jeden Hermiteův interpolační polynom H funkce f v uzlech $x_0, x_1, \ldots, x_n$.

KONSTRUKCE HERMITEOVA INTERPOLAČNÍHO POLYNOMU

I. V základním tvaru: Podle Věty 4 je obecný tvar polynomu

$$H(x) = a_0 + a_1 x + \ldots + a_{2n+1} x^{2n+1}.$$

Koeficienty $a_0, a_1, \ldots, a_{2n+1}$ se volí tak, aby polynom H splňoval podmínky (40).

Příklad 1. Najděte Hermiteův interpolační polynom funkce $f(x)$ v uzlech z tabulky.

k	x_k	$f(x_k)$	$f'(x_k)$
0	-1	2	1
1	1	0	1

$$H(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \qquad H'(x) = a_1 + 2a_2 x + 3a_3 x^2.$$

Dosazením $x_0 = -1$ a $x_1 = 1$ za x vznikne systém rovnic:

$$\begin{aligned} a_0 - a_1 + a_2 - a_3 &= 2 \\ a_1 - 2a_2 + 3a_3 &= 1 \\ a_0 + a_1 + a_2 + a_3 &= 0 \\ a_1 + 2a_2 + 3a_3 &= 1 \end{aligned}$$

Řešením tohoto systému rovnic dostaneme koeficienty $a_0 = 1$, $a_1 = -2$, $a_2 = 0$, $a_3 = 1$, takže Hermiteův interpolační polynom je $H(x) = 1 - 2x + x^3$.

II. V zobecněném Newtonově tvaru: Oproti úloze Lagrangeovy interpolace jsou zde v každém uzlu dány dvě hodnoty. Zapíšeme tedy každý uzel do tabulky dvakrát. Vznikne potřeba počítat poměrnou diferenci $f[x_i, x_i]$, která nemá smysl. Je však přirozené ji definovat předpisem

$$f[x_i, x_i] = \lim_{x \longrightarrow x_i} f[x_i, x] = \lim_{x \longrightarrow x_i} \frac{f(x) - f(x_i)}{x - x_i} = f'(x_i)$$

Při výpočtu poměrných diferencí vyšších řádů již nevznikají žádné potíže. Tedy Hermiteův interpolační polynom $H(x)$ má tvar

$$\begin{aligned} H(x) &= f_0 + f[x_0, x_0](x - x_0) + f[x_0, x_0, x_1](x - x_0)^2 + \ldots \\ &+ f[x_0, x_0, \ldots, x_n, x_n](x - x_0)^2 \ldots (x - x_{n-1})^2 (x - x_n) \end{aligned}$$

Příklad 2. Úlohu z příkladu 1 řešte pomocí zobecněného Newtonova polynomu.

i	x_i	f_i	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$
0	-1	2	1	-1	1
1	-1	2	-1	1	
2	1	0	1		
3	1	0			

Tedy $H(x) = 2 + (x+1) - (x+1)^2 + (x+1)^2(x-1) = x^3 - 2x + 1$.

Příklad 3. Najděte Hermiteův interpolační polynom funkce f v uzlech z tabulky.

i	x_i	f_i	f_i'
0	-1	1	2
1	0	2	1
2	0,5	1	-1

Řešení:

i	x_i	f_i	$f[x_i, x_{i+1}]$				
0	-1	1	2	-1	1	$-\frac{10}{3}$	$\frac{100}{9}$
1	-1	1	1	0	-4	$\frac{40}{3}$	
2	0	2	1	-6	16		
3	0	2	-2	2			
4	0,5	1	-1				
5	0,5	1					

Tedy

$$\begin{aligned} H(x) &= 1 + 2(x+1) - (x+1)^2 + (x+1)^2 x - \frac{10}{3}(x+1)^2 x^2 \\ &+ \frac{100}{9}(x+1)^2 x^2 (x - 0,5). \end{aligned}$$

(B) φ je interpolační Hermiteův kubický splajn.

I v případě Hermiteovy interpolace je nevýhodné pracovat s polynomy vysokých stupňů. Pro řešení ÚLOHY HERMITEOVY INTERPOLACE se velmi osvědčuje tento typ po částech polynomiální funkce:

Definice. Nechť $a = x_0 < x_1 < \ldots < x_n = b$. Hermiteův kubický splajn s uzly $x_0, x_1, \ldots, x_n$ je každá funkce $s(x)$ na $\langle a, b\rangle$ s těmito vlastnostmi (a), (b):

(a) Na intervalu $\langle x_{i-1}, x_i\rangle$ je $s(x)$ polynom nejvýše třetího stupně (tento polynom označíme $s_i(x)$) pro $i = 1, 2, \ldots, n$

(b) $s(x) \in C^{(1)}\langle a, b\rangle$, tj. s a s' jsou spojité na $\langle a, b\rangle$.

Všimněte si, že Hermiteův kubický splajn obecně není kubický splajn.

Definice. Nechť $a = x_0 < x_1 < \ldots < x_n = b$ a $f \in C^{(1)}\langle a, b\rangle$. Hermiteův kubický splajn $s(x)$ s uzly $x_0, x_1, \ldots, x_n$ splňující

$$s(x_i) = f(x_i) \quad \text{a} \quad s'(x_i) = f'(x_i)$$

pro $i = 0, 1, \ldots, n$ se nazývá *interpolační Hermiteův kubický splajn* funkce f v uzlech $x_0, x_1, \ldots, x_n$.

Podle Věty 4 je polynom $s_i(x)$ pro $i = 1, 2, \ldots, n$ jednoznačně určen požadavky

$$s_i(x_j) = f(x_j) \quad \text{a} \quad s_i'(x_j) = f'(x_j) \quad pro \quad j = i-1, i.$$

Tento polynom lze zkonstruovat konstrukcí I nebo II.

Příklad 4. Najděte interpolační Hermiteův kubický splajn funkce f v uzlech x_0, x_1, x_2 z příkladu 3.

1. Interval $\langle x_0, x_1\rangle = \langle -1, 0\rangle$:

i	x_i	f_i	$f[x_i, x_{i+1}]$		
0	-1	1	2	-1	1
1	-1	1	1	0	
3	0	2	1		
4	0	2			

2. Interval $\langle x_1, x_2\rangle = \langle 0, 0, 5\rangle$:

i	x_i	f_i	$f[x_i, x_{i+1}]$		
0	0	2	1	-6	16
1	0	2	-2	2	
3	0,5	1	-1		
4	0,5	1			

Tedy $s(x) = \begin{cases} s_1(x) = 1 + 2(x+1) - (x+1)^2 + (x+1)^2 x & pro \quad -1 \le x \le 0 \\ s_2(x) = 2 + x - 6x^2 + 16x^2(x - 0,5) & pro \quad 0 \le x \le 0,5 \end{cases}$

12.4 Aproximace diskrétní metodou nejmenších čtverců

Viz skripta Dalík: Numerické metody (stručně skripta) odst. 10.5, Úloha 2.

12.5 Aproximace funkce vyrovnávacími kubickými splajny

ÚLOHA APROXIMACE (2): Nechť $f \in C\langle a, b\rangle$, $a = x_0 < x_1 < \ldots < x_n = b$ jsou uzly, $f_0 = f(x_0), \ldots, f_n = f(x_n)$ a $p_0, p_1, \ldots, p_n$ jsou kladná čísla (váhy). Najděte funkci $u \in C^{(2)}\langle a, b\rangle$ splňující

$$\Phi(u) \le \Phi(h) \quad pro\ vechna \quad h \in C^{(2)}\langle a, b\rangle. \tag{41}$$

Zde $\Phi(h) = \int_a^b h''^2 dx + \sum_{k=0}^n p_k(h(x_k) - f_k)^2$.

Věta 5. Úloha aproximace má jediné řešení, kterým je kubický splajn příslušný uzlům $x_0, x_1, \ldots, x_n$.

Důkaz. Nechť $u_0 \in C^{(2)}\langle a, b\rangle$ splňuje (41). Sestrojme kubický splajn $g(x)$ příslušný uzlům $x_0, \ldots, x_n$ tak, aby $g(x_k) = u_0(x_k)$ pro $k = 0, \ldots, n$. Druhé sčítance ve výrazech $\Phi(u_0)$ a $\Phi(g)$ jsou stejné a tedy podle Věty 3 platí $\Phi(g) \leq \Phi(u_0)$. Zároveň však $\Phi(u_0) \leq \Phi(g)$ podle (41) a tedy $\Phi(g) \leq \Phi(u_0)$. Odtud plyne $u_0 = g$ podle Věty 3.

KONSTRUKCE VYROVNÁVACÍHO KUBICKÉHO SPLAJNU: Hledáme kubický splajn g tak, aby výraz

$$\Phi(g) = \int_a^b g''^2 dx + \sum_{k=0}^n p_k(g_k - f_k)^2 = F(g_0, g_1, \ldots, g_n),$$

kde $g_i = g(x_i)$ pro $i = 0, 1, \ldots, n$, byl minimální. Nutnou podmínkou minima je

$$\frac{\partial F}{\partial g_s} = 0 \qquad pro \qquad s = 0, 1, \ldots, n \Longleftrightarrow$$
$$(A + H\,P^{-1}H^\top)\vec{m} = H\,\vec{f} \qquad a \qquad \vec{g} = \vec{f} - P^{-1}H^\top\vec{m}.$$

Zde A, H, $\vec{m}$ a $\vec{f}$ jsou definovány v odst. 12.2 a $P = diag(p_0, p_1, \ldots, p_n)$.

13 Numerické derivování a numerická integrace

13.1 Numerické derivování

Viz skripta odst. 11.1

13.2 Numerická integrace

Viz skripta odst. 11.3-11.5 bez Rombergovy metody.