Obsah

Předmluva

Současná doba je charakterizována prudkým rozvojem výpočetní techniky a s tím souvisí rozšíření možností aplikace matematiky i v dalších vědeckých oborech — biologii, chemii, ekonomii, psychologii, lékařství a v technických vědách.

Důležitou úlohu v řadě aplikací mají metody numerické matematiky a odpovídající efektivní algoritmy. A právě základním numerickým metodám jsou věnována tato skripta.

Skripta vycházejí ve druhém, rozšířeném vydání. Svým rozsahem pokrývají dvousemestrovou přednášku z numerických metod v rámci studijních programů Matematika a Aplikovaná matematika na Přírodovědecké fakultě MU.

Skripta jsou věnována základním numerickým metodám, a protože odpovídající algoritmy pro realizaci těchto metod jsou poměrně jednoduché, nejsou, až na výjimky, v těchto skriptech uvedeny. Konstrukci těchto algoritmů v rámci systému MATLAB je věnováno dost prostoru v příslušných cvičeních k uvedeným přednáškám.

Brno, říjen 2008

Ivana Horová
Jiří Zelinka

Úvod

Hamming, R. W. (1962):
„*Cíl výpočtů – pochopení podstaty, a ne číslo*“
„*Dříve než budete úlohu řešit, promyslete si, co budete dělat s jejím řešením.*“

Numerická matematika se zabývá procesy, pomocí nichž lze matematické problémy řešit aritmetickými operacemi. Někdy to bude znamenat sestrojení algoritmů k řešení problému, který je již v takovém tvaru, že jeho řešení lze nalézt aritmetickými prostředky (např. systém lineárních rovnic). Často to bude znamenat náhradu veličin, které nemohou být počítány aritmeticky (např. derivace nebo integrály) aproximacemi, které umožní nalézt přibližné řešení. Numerická matematika se rovněž zabývá volbou postupu, který se „nejlépe“ hodí k řešení speciálního problému. Uvádíme proto řadu příkladů na ilustraci numerických metod. Účelem těchto příkladů je, aby pomohly čtenáři porozumět podstatě té které numerické metody.

Poznámka 1. V 9. stol. n. l. arabský matematik Muhamad ibn Músá al-Chvarízmí napsal knihu, ve které vykládá indický početní systém. Latinskému překladu názvu knihy „Algorithmi de numero Indorum“ vděčíme za název algoritmus.

Numerická řešení problémů jsou obvykle zatížena chybami, které vznikají ve dvou oblastech: těmi, které jsou obsaženy v matematické formulaci problému, a těmi, které jsou způsobeny hledáním řešení numerickou cestou.

První kategorie zahrnuje chyby způsobené tím, že matematický problém je pouze aproximací reálné situace. Jiným pramenem chyb jsou např. nepřesnosti fyzikálních konstant nebo chyby v empirických hodnotách.

Nechť x je přesné číslo a nechť $\tilde{x}$ značí aproximaci x. Rozdíl $\tilde{x} - x$ nazýváme *absolutní chybou* aproximace $\tilde{x}$, veličinu α, $|\tilde{x} - x| \leq \alpha$, nazýváme *odhadem absolutní chyby*. Vhodněji lze vyjádřit vztah mezi x a $\tilde{x}$ prostřednictvím relativní chyby:

Podíl $(x - \tilde{x})/x$ nazýváme *relativní chybou* a veličinu δ,

$$\left| \frac{x - \tilde{x}}{x} \right| \leq \delta,$$

nazýváme *odhadem relativní chyby*.

Jestliže $|x|$ je malé číslo, je vhodné pro odhad chyby použít relativní chyby, což je vidět z následujícího příkladu.

Mějme např. čísla $x_1 = 1{,}31$, $\tilde{x}_1 = 1{,}30$ a $x_2 = 0{,}12$, $\tilde{x}_2 = 0{,}11$. Pro absolutní chyby v obou případech platí

$$|x_1 - \tilde{x}_1| = 0{,}01, \qquad |x_2 - \tilde{x}_2| = 0{,}01,$$

ale pro relativní chyby platí

$$\left|\frac{\tilde{x}_1 - x_1}{x_1}\right| = 0{,}0076, \qquad \left|\frac{\tilde{x}_2 - x_2}{x_2}\right| = 0{,}0833.$$

Tento výsledek ukazuje, že $\tilde{x}_1$ je bližší x_1 než $\tilde{x}_2$ k x_2, zatímco z absolutní chyby nic takového neplyne.

Relativní chyba rovněž slouží k odhadu platných cifer aproximace $\tilde{x}$. Tuto skutečnost lze formulovat takto:

Definice 0.1. Řekneme, že aproximace $\tilde{x}$ čísla x má s platných cifer, jestliže s je největší celé nezáporné číslo takové, že platí

$$\left|\frac{x - \tilde{x}}{x}\right| \leq 5.10^{-s}.$$

Poznámka 2. Nechť x je reálné číslo, které má obecně nekonečné dekadické vyjádření. Číslo $x^{(d)}$, které má d desetinných míst, je správně zaokrouhlenou hodnotou čísla x, platí-li

$$|x - x^{(d)}| \leq \frac{1}{2}10^{-d}.$$

Ve správně zaokrouhleném čísle jsou všechny cifry platné.

Dalším zdrojem chyb během výpočtu je nepřesné zobrazování čísel v paměti počítače jako důsledek její konečné velikosti.

§ 0.1. Reprezentace čísel v počítači[1]

1. ***Strojová reprezentace celých čísel na*** n ***bitů (počítání modulo*** 2^n***)***

(i) *bez znaménka (unsigned integer)*:

$$a \geq 0 \quad \Rightarrow \quad \text{rozsah: } 0 \leq a \leq 2^n - 1 = (\underbrace{11\ldots1}_{n})_2,$$

(ii) *se znaménkem (signed integer)*:

$$a \text{ libovolné} \quad \Rightarrow \quad \text{rozsah: } (\underbrace{10\ldots0}_{n})_2 = -2^{n-1} \leq a \leq 2^{n-1} - 1 = (\underbrace{01\ldots1}_{n})_2.$$

Současně vidíme, že prvý bit určuje znaménko: 1 = minus, 0 = plus.

[1] Tato část byla převzata z nepublikovaného učebního textu doc. RNDr. V. Veselého, CSc.

Je zřejmé, že *v rámci uvedených rozsahů jsou celočíselné výpočty absolutně přesné*, zatímco *mimo ně naopak zcela chybné*.

Příklad 0.1. ($n = 3$)

(i) $0 \le a \le 7$: $\overset{\text{mod8}}{\boxed{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7}}\ 0\ 1\ 2\ 3\ldots$

(ii) $-4 \le a \le 3$: $0\ -7\ -6\ -5\ \overset{\text{mod8}}{\boxed{-4\ -3\ -2\ -1\ 0\ 1\ 2\ 3}}\ldots$

K číslu a najdeme v modulární aritmetice číslo opačné $-a$ snadno z rovnice

$$\underbrace{a + \bar{a}}_{(11\ldots1)_2} + 1 = 2^n \equiv 0 \bmod 2^n.$$

Odtud dostáváme $-a = \bar{a} + 1$, kde $\bar{a}$ vznikne z a negací po bitech.

Například

$$\begin{array}{rcl} a = 2 & = & (010)_2 \\ \bar{2} & = & (101)_2 \\ \hline 2 + \bar{2} & = & (111)_2 \\ & \Downarrow & \\ -2 = \bar{2} + 1 & = & (110)_2 \\ & = & 6 \bmod 8. \end{array}$$

V počítači se celá čísla zobrazují zpravidla v těchto přesnostech:

$n = 8$	(1 bajt)	...	(un)signed char (1 znak)
$n = 16$	(2 bajty)	...	(un)signed short integer (poloviční přesnost)
$n = 32$	(4 bajty)	...	(un)signed integer (jednoduchá přesnost)
$n = 64$	(8 bajtů)	...	(un)signed long integer (dvojnásobná přesnost)

2. *Strojová reprezentace reálných čísel na n bitů*

Nechť $q \ge 2$ značí základ číselné soustavy. V počítačích pracujeme s čísly nejčastěji v soustavách $q = 2, 8, 16$. Přesná reálná čísla se reprezentují v tzv. *semilogaritmickém tvaru pohyblivé řádové čárky* (normalizovaná mantisa + exponent):

$$p = \overbrace{\pm d_1, d_2 d_3 \ldots d_k d_{k+1}}^{\text{mantisa}} \ldots \times q^e,$$

kde $e \in \mathbb{Z}$ je exponent a $1 \le d_1 \le q-1$, $0 \le d_j \le q-1$ (pro $j > 1$) jsou cifry mantisy. Zejména v případě $q = 2$ je $d_1 = 1$, takže tento bit je možno využít pro zobrazení znaménka mantisy (1 = minus, 0 = plus).

Strojová reálná čísla se ukládají pouze s konečným počtem k cifer mantisy. Obdržíme tak přibližnou reprezentaci $\mathrm{fl}(p)$ (floating-point representation), která vznikne buď pouhým *odseknutím* (*chopping*) přebývajících cifer nebo se navíc poslední k-tá cifra d_k *zaokrouhlí* (*rounding*). Současně se také vhodně *omezí rozsah exponentu*: $-e_{\min} \le e \le e_{\max}$. Dostáváme tedy tyto aproximace:

a) $\tilde{p} = \mathrm{fl}_{\mathrm{chop}}(p) = \pm d_1, d_2 d_3 \ldots d_k \times q^e$,
s absolutní chybou aproximace $0 \le |p - \tilde{p}| < q^{e-(k-1)}$,

b) $\tilde{p} = \mathrm{fl}_{\mathrm{round}}(p) = \pm d_1, d_2 d_3 \ldots d_{k-1} \tilde{d}_k \times q^e$, $\tilde{d}_k = \mathrm{round}(d_k, d_{k+1}, \ldots)$,
s absolutní chybou aproximace $0 \le |p - \tilde{p}| \le q^{e-(k-1)}/2$.

Příklad 0.2.

(a) $q = 10$, $k = 6$:

$$p = \tfrac{22}{7} = 3{,}1428\overline{571428} \quad \Rightarrow \quad \begin{cases} \mathrm{fl}_{\mathrm{chop}}(p) &= 3{,}14285 \times 10^0 \\ \mathrm{fl}_{\mathrm{round}}(p) &= 3{,}14286 \times 10^0 \end{cases}$$

(b) $q = 2$, $k = 5$:

$$p = 0{,}1 = (1{,}1001\overline{1001})_2 \times 2^{-4} \quad \Rightarrow \quad \begin{cases} \mathrm{fl}_{\mathrm{chop}}(p) &= (1{,}1001)_2 \times 2^{-4} \\ \mathrm{fl}_{\mathrm{round}}(p) &= (1{,}1010)_2 \times 2^{-4} \end{cases}$$

Tento příklad je současně ilustrací čísla, které má *konečný počet cifer v dekadické soustavě, ale nekonečný počet cifer v binární soustavě* a *není* tedy v paměti počítače zobrazeno přesně.

V počítačích se reálná čísla zobrazují zpravidla v těchto přesnostech:

a) *Jednoduchá přesnost (4 bajty):* $n = 32 = 24$ bitů mantisy + 8 bitů exponentu.
Rozsah exponentu: $\underbrace{-2^7}_{-128} \le e \le \underbrace{2^7 - 1}_{127}$.
Dekadický rozsah:
$2{,}938736 \times 10^{-39}$ až $1{,}701412 \times 10^{38}$, kde
$2{,}938736 \times 10^{-39} \doteq 1 \times 2^{-128}$ a
$1{,}701412 \times 10^{38} \doteq \underbrace{(1{,}11\ldots1)_2}_{\approx 2} \times 2^{127} \doteq 2 \times 2^{127} = 2^{128}$.
Dekadická přesnost mantisy:
$2^{-23} \doteq 1{,}2 \times 10^{-7} \Rightarrow$ *cca 7 dekadických cifer přesnosti*, což však vzhledem k příkladu 0.2(b) neznamená, že každé číslo s nejvýše 7 dekadickými ciframi musí být zobrazeno přesně.

b) *Dvojnásobná přesnost (8 bajtů):* $n = 64 = 53$ bitů mantisy + 11 bitů exponentu.
Rozsah exponentu: $\underbrace{-2^{10}}_{-1024} \le e \le \underbrace{2^{10} - 1}_{1023}$.
Dekadický rozsah:
$5{,}562684646268003 \times 10^{-309}$ až $8{,}988465674311580 \times 10^{307}$, kde
$5{,}562684646268003 \times 10^{-309} \doteq 1 \times 2^{-1024}$ a
$8{,}988465674311580 \times 10^{307} \doteq \underbrace{(1{,}11\ldots1)_2}_{\approx 2} \times 2^{1023} \doteq 2 \times 2^{1023} = 2^{1024}$.
Dekadická přesnost mantisy:
$2^{-52} \doteq 2{,}2 \times 10^{-16} \Rightarrow$ *cca 16 dekadických cifer přesnosti.*

Běžně používaná binární reprezentace dle normy IEEE (např. počítače třídy PC) je v poněkud modifikovaném tvaru:

$$\tilde{p} = \mathrm{fl}_{\mathrm{IEEE}}(p) = s\,\bar{e}_{10}e_9e_8\ldots e_0 d_2, d_3\ldots d_{53},$$

kde

- $s1\ d_2, d_3\ldots d_{53}$ ... mantisa se znaménkovým bitem s (1=minus, 0=plus) a dvěma binárními místy před řádovou čárkou ($d_1 = 1$ a d_2),
- $e = (\bar{e}_{10}e_9e_8\ldots e_0)_2$ ... exponent v 11-bitové binární reprezentaci se znaménkem podle 1ii v symetrickém rozsahu $-(2^{10}-1) \le e \le 2^{10}-1$. Tedy $\bar{e}_{10} = 1$ odpovídá nezaporným hodnotám exponentu a $\bar{e}_{10} = 0$ záporným hodnotám, přičemž případ $e = -2^{10}$ byl vyloučen, neboť pro něj jsou všechny bity exponentu $\bar{e}_{10}e_9e_8\ldots e_0$ nulové a vznikla by kolize s vyjádřením čísla 0, která je dána nulovými hodnotami všech bitů IEEE reprezentace (jinak by totiž tyto nulové bity určovaly kladné číslo $(10{,}0\ldots0)_2 \times 2^{-2^{10}}$).

Zmíníme se ještě o šíření chyb při provádění aritmetických operací v případě zápisu čísel v pohyblivé řádové čárce. Lze dokázat, že platí ([6],[13])

$$\mathrm{fl}(x * y) = (x * y)(1 + \delta),$$

kde $*$ znamená libovolnou z aritmetických operací $+, -, \times, :$, a $|\delta| \le \mu$, μ dekadická přesnost mantisy.

§ 0.2. Celková chyba výpočtu

Dalším zdrojem chyb je skutečnost, že se neřeší problém, který byl původně zadán, ale nějaká jeho aproximace. Často je to způsobeno náhradou procesu nekonečného procesem konečným. Přesněji: Předpokládejme, že veličina Y je jednoznačně určena hodnotami $x_1, \ldots, x_n$, tj.

$$Y = F(x_1, \ldots, x_n).$$

Funkční závislost F nahradíme numerickou metodou f a získané teoretické řešení označíme y:

$$y = f(x_1, \ldots, x_n).$$

Vzhledem k tomu, že místo hodnot x_i musíme často používat jen aproximace $\tilde{x}_i$, a protože nelze provádět všechny aritmetické operace přesně (zaokrouhlování mezivýsledků), bude se vypočtená hodnota

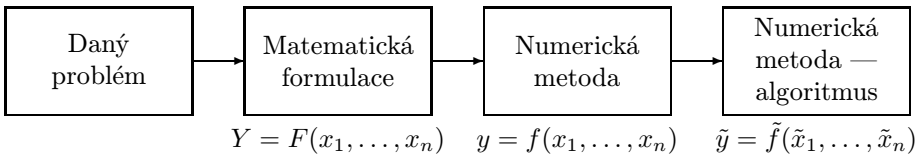
$$\tilde{y} = \tilde{f}(\tilde{x}_1, \ldots, \tilde{x}_n)$$

lišit od y. Celkovou chybu vyjádříme jako součet dílčích chyb takto:

$$\begin{aligned} Y - \tilde{y} = {} & \{Y - y\} + \{f(x_1, x_2, \ldots, x_n) - f(\tilde{x}_1, \ldots, \tilde{x}_n)\} + \\ & + \left\{f(\tilde{x}_1, \ldots, \tilde{x}_n) - \tilde{f}(\tilde{x}_1, \ldots, \tilde{x}_n)\right\}. \end{aligned}$$

Rozdíl $Y - y$ nazveme *chybou metody*, $f(x_1, x_2, \ldots, x_n) - f(\tilde{x}_1, \ldots, \tilde{x}_n)$ se nazývá *chyba primární* a $f(\tilde{x}_1, \ldots, \tilde{x}_n) - \tilde{f}(\tilde{x}_1, \ldots, \tilde{x}_n)$ se nazývá *chyba sekundární*.

Schematicky lze předchozí úvahy znázornit takto:



U každé numerické metody by měla být uvedena chyba této metody. Primární chybu lze odhadnout následujícím způsobem:

Věta 0.1. *Buď $U = \{x_i : |x_i - \tilde{x}_i| \leq \alpha_i, i = 1, \ldots, n\}$ a nechť funkce $f(x_1, \ldots, x_n)$ je spojitě diferencovatelná na U. Pak*

$$|f(x_1, \ldots, x_n) - f(\tilde{x}_1, \ldots, \tilde{x}_n)| \leq \sum_{i=1}^{n} A_i \alpha_i, \tag{0.1}$$

kde

$$A_i = \sup_U \left| \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) \right|, \qquad i = 1, \ldots, n.$$

Důkaz plyne z Lagrangeovy věty pro funkce n proměnných.

Poznámka 3. V praxi se užívá odhadu

$$|f(x_1, \ldots, x_n) - f(\tilde{x}_1, \ldots, \tilde{x}_n)| \leq \sum_{i=1}^{n} \left| \frac{\partial f}{\partial x_i}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \right| \alpha_i. \tag{0.2}$$

Odhad sekundární chyby lze provést teprve tehdy, až je algoritmus rozepsán do posloupnosti aritmetických operací (viz příklad 0.4).

Příklad 0.3. Počítejme gravitační zrychlení g ze vzorce pro dobu kmitu T matematického kyvadla

$$T = 2\pi \sqrt{\frac{l}{g}}$$

(l délka kyvadla). Jaké absolutní chyby se dopustíme, jestliže dobu kmitu T měříme s chybou ΔT a l s chybou Δl?

Řešení. Je třeba v podstatě odhadnout primární chybu:

$$|g(l, T) - g(\tilde{l}, \tilde{T})|,$$

kde $\tilde{T} = T + \Delta T$, $\tilde{l} = l + \Delta l$, a funkce g je dána vztahem

$$g(l, T) = \frac{4\pi^2 l}{T^2}.$$

Počítejme parciální derivace funkce g podle jednotlivých proměnných:

$$\frac{\partial g}{\partial l} = \frac{4\pi^2}{T^2}, \qquad \frac{\partial g}{\partial T} = -\frac{8\pi^2 l}{T^3}$$

Užijeme vztahu (0.2):

$$\begin{aligned} |g(l,T) - g(\tilde{l},\tilde{T})| &\le \left|\frac{\partial g}{\partial l}(\tilde{l},\tilde{T})\right| |\Delta l| + \left|\frac{\partial g}{\partial T}(\tilde{l},\tilde{T})\right| |\Delta T| = \\ &= \frac{4\pi^2}{\tilde{T}^2}|\Delta l| + \frac{8\pi^2 l}{\tilde{T}^3}|\Delta T| = \frac{4\pi^2}{\tilde{T}^2}\tilde{l}\left(\frac{|\Delta l|}{\tilde{l}} + 2\frac{|\Delta T|}{\tilde{T}}\right) = \\ &= g(\tilde{l},\tilde{T})\left(\frac{|\Delta l|}{\tilde{l}} + 2\frac{|\Delta T|}{\tilde{T}}\right). \end{aligned}$$

Ukážeme nyní na příkladě, že sekundární chyba a tedy i celková chyba výpočtu závisí na tom, jak výpočet uspořádáme a jak zaokrouhlujeme během výpočtového procesu.

Příklad 0.4. Nechť $\boldsymbol{u} = (u_1, \dots, u_n)$, $\boldsymbol{v} = (v_1, \dots, v_n)$ jsou dva vektory, jejichž všechny složky jsou správně zaokrouhleny na d desetinných míst a nechť $\boldsymbol{U} = (U_1, \dots, U_n)$, $\boldsymbol{V} = (V_1, \dots, V_n)$ jsou přesné vektory. Je třeba vypočítat skalární součin

$$(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} u_i v_i$$

na d míst.
Ptáme se: Je výhodnější každý ze součinů $u_i v_i$ napřed zaokrouhlit na d míst a pak sečíst, nebo napřed všechny součiny $u_i v_i$ sečíst a pak zaokrouhlit výsledek na d desetinných míst?
Řešení. Nechť

$$U_i = u_i + \varepsilon_i, \quad V_i = v_i + \delta_i, \quad i = 1, \dots, n, \quad |\varepsilon_i|,\ |\delta_i| \le \frac{1}{2}10^{-d}.$$

Pak

$$u_i v_i = (U_i - \varepsilon_i)(V_i - \delta_i) = U_i V_i - \varepsilon_i V_i - \delta_i U_i + \varepsilon_i \delta_i\,. \tag{0.3}$$

Vztah (0.3) udává primární chybu $U_i V_i - u_i v_i$. Jestliže každý ze součinů $u_i v_i$ zaokrouhlíme na d desetinných míst a výsledek označíme $(u_i v_i)_z$, máme

$$(u_i v_i)_z = U_i V_i - \varepsilon_i V_i - \delta_i U_i + \varepsilon_i \delta_i + \gamma_i, \tag{0.4}$$

kde $|\gamma_i| \le \frac{1}{2}10^{-d}$. V prvním případě je skalární součin $(\boldsymbol{u}, \boldsymbol{v})$ roven

$$(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} (u_i v_i)_z = \sum_{i=1}^{n} U_i V_i - \sum_{i=1}^{n} (\delta_i U_i + \varepsilon_i V_i - \varepsilon_i \delta_i - \gamma_i),$$

takže celková chyba je

$$\begin{aligned}(\boldsymbol{U},\boldsymbol{V})-(\boldsymbol{u},\boldsymbol{v}) &= \sum_{i=1}^{n}(\delta_i U_i+\varepsilon_i V_i-\varepsilon_i\delta_i-\gamma_i) = \\ &= \sum_{i=1}^{n}(\delta_i U_i+\varepsilon_i V_i-\varepsilon_i\delta_i)-\sum_{i=1}^{n}\gamma_i\ .\end{aligned} \tag{0.5}$$

Při druhém postupu počítáme bez zaokrouhlování, tj. každý součin $u_i v_i$ má $2d$ desetinných míst. Tyto součiny sečteme a výsledek zaokrouhlíme:

$$(\boldsymbol{u},\boldsymbol{v})=\sum_{i=1}^{n}u_i v_i+\gamma=\sum_{i=1}^{n}U_i V_i-\sum_{i=1}^{n}(\delta_i U_i+\varepsilon_i V_i-\delta_i\varepsilon_i)+\gamma, \tag{0.6}$$

kde $|\gamma|\le\frac{1}{2}10^{-d}$. Tedy

$$(\boldsymbol{U},\boldsymbol{V})-(\boldsymbol{u},\boldsymbol{v})=\sum_{i=1}^{n}(\delta_i U_i+\varepsilon_i V_i-\delta_i\varepsilon_i)-\gamma\ . \tag{0.7}$$

Protože $u_i v_i$ jsou správně zaokrouhlená čísla na d desetinných míst, je

$$\left|\sum_{i=1}^{n}(\delta_i U_i+\varepsilon_i V_i-\delta_i\varepsilon_i)\right|\le\frac{1}{2}10^{-d}\sum_{i=1}^{n}(|U_i|+|V_i|+\frac{1}{2}10^{-d})\ . \tag{0.8}$$

Dále je

$$\sum_{i=1}^{n}|\gamma_i|\le\frac{n}{2}10^{-d}\ .$$

V rovnicích (0.5) a (0.7) je pouze poslední člen způsoben zaokrouhlovacími chybami během výpočtu. Nechť $|U_i|$, $|V_i|<1$, pro $i=1,\dots,n$. Pak druhý člen v rovnici (0.5) je řádově stejný jako první člen, ale druhý člen v rovnici (0.7) je malý ve srovnání s prvním členem. Druhý postup je proto výhodnější než první postup. Tyto úvahy mají velký význam při výpočtech na počítačích, kde je velký počet operací, a pořadí, v němž je výpočet proveden, je velmi významným faktorem.

§ 0.3. Podmíněnost úloh

Předpokládejme, že B_1 (množina vstupních dat) a B_2 (množina výstupních dat) jsou Banachovy prostory.

Řekneme, že úloha

$$y=U(x),\qquad x\in B_1,\qquad y\in B_2$$

je *korektní* pro dvojici prostorů (B_1,B_2), jestliže

1. ke každému $x\in B_1$ existuje jediné řešení $y\in B_2\colon y=U(x)$.

2. toto řešení spojitě závisí na vstupních datech, tj. jestliže $x_n \to x$, $U(x_n) = y_n$, pak $U(x_n) \to U(x) = y$.

Velkou třídu nekorektních úloh tvoří nejednoznačně řešitelné úlohy.

Uvedeme nyní charakteristiku dobře podmíněných úloh. Řekneme, že korektní úloha je *dobře podmíněna*, jestliže malá změna ve vstupních datech vyvolá malou změnu řešení. Je-li $y + \Delta y$ resp. y řešení odpovídající vstupním datům $x + \Delta x$ resp. x, potom číslo

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

(kde místo absolutních hodnot mohou být obecně normy, viz kap. 1) nazýváme *číslem podmíněnosti* úlohy $y = U(x)$. Je-li $C_p \approx 1$, je úloha velmi dobře podmíněna. Pro velká C_p (> 100) je úloha špatně podmíněna.

Posuďme nyní z hlediska dobré či špatné podmíněnosti výpočet hodnoty $y = \sin x$.

Číslo podmíněnosti je v tomto případě dáno vztahem

$$C_p = \frac{\left|\dfrac{\Delta y}{y}\right|}{\left|\dfrac{\Delta x}{x}\right|} = \frac{\left|\dfrac{\sin(x+\Delta x) - \sin x}{\sin x}\right|}{\left|\dfrac{\Delta x}{x}\right|} = \left|\frac{\sin(x+\Delta x) - \sin x}{\Delta x}\right| \left|\frac{x}{\sin x}\right|.$$

Nechť $\Delta x \to 0$ a zabývejme se výpočtem $\sin x$ a) v okolí bodu 0, b) v okolí bodu π.

a) V okolí bodu 0 je
$$C_p \approx |\cos x| \left|\frac{x}{\sin x}\right| \approx 1.$$

b) V okolí bodu π
$$C_p = \left|\frac{x \cos x}{\sin x}\right| = |x \operatorname{cotg} x| \to +\infty.$$

Speciálně pro $x = 3{,}14$, $\Delta x = 0{,}01$ se dá ukázat ([18])

$$C_p \approx 1972.$$

Úloha stanovit $\sin x$ *v okolí 0 je dobře podmíněna* a *v okolí bodu* π *je špatně podmíněna*.

§ 0.4. Realizace numerických výpočtů

Následující příklady ukazují na problémy, které se mohou objevit při realizaci numerických výpočtů.

Příklad 0.5. Počítejme rozdíl dvou čísel $x = 0{,}54617$ a $y = 0{,}54601$. Přesná hodnota rozdílu je $d = 0{,}00016$. Uvažujme nyní čísla zaokrouhlená:

$$\tilde{x} = 0{,}5462, \qquad \tilde{y} = 0{,}5460.$$

Nyní je $\tilde{d} = \tilde{x} - \tilde{y} = 0{,}0002$. Relativní chyba je v tomto případě

$$\frac{|d - \tilde{d}|}{|d|} = 0{,}25$$

a je tedy dosti velká.

Co se zde stalo? Čísla $\tilde{x}$ a $\tilde{y}$ jsou „téměř“ stejná (při zaokrouhlení na 4 cifry). Při odčítání platné cifry se vyruší a zůstanou „méně“ významné cifry. Tento jev se nazývá „katastrofické zrušení“ a vyskytuje se v případech, kdy odečítáme přibližně dvě stejná čísla. Ale při konkrétních výpočtech můžeme tento jev eliminovat.

Uvažujme například kvadratickou rovnici

$$ax^2 + bx + c = 0, \qquad a \neq 0.$$

Kořeny této rovnice jsou

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \qquad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Je jasné, že v případě, kdy $-b$ a $\sqrt{b^2 - 4ac}$ jsou „blízká“ čísla, výpočet x_2 může být velmi negativně ovlivněn „katastrofickým zrušením“. Tomuto problému se můžeme vyhnout tak, že počítáme kořeny následujícím způsobem:

$$x_1 = \frac{-b - \operatorname{sign}(b)\sqrt{b^2 - 4ac}}{2a}, \qquad x_2 = \frac{c}{ax_1}.$$

Příklad 0.6. Počítejme integrál

$$E_n = \int_0^1 x^n \mathrm{e}^{x-1} \,\mathrm{d}x$$

pro $n = 1, 2, 3, \ldots$

Integrací per partes dostaneme

$$E_n = \int_0^1 x^n \mathrm{e}^{x-1} \,\mathrm{d}x = \left[x^n \mathrm{e}^{x-1}\right]_0^1 - \int_0^1 n x^{n-1} \mathrm{e}^{x-1} \,\mathrm{d}x,$$

neboli

$$E_n = 1 - nE_{n-1}.$$

Jelikož $E_1 = 1/\mathrm{e}$, lze užitím této formule vypočítat E_n pro $n = 2, 3, \ldots$

Nechť $E_1 = 0{,}367879$ (tj. hodnota 1/e je zaokrouhlena na 6 cifer). Pak

$$\begin{aligned} E_2 &= 0{,}264242 \\ E_3 &= 0{,}207274 \\ &\vdots \\ E_9 &= -0{,}0684800 \end{aligned}$$

I když je integrand kladný, je hodnota integrálu záporná! Tento jev můžeme vysvětlit takto: Ze vztahu $E_2 = 1 - 2E_1$ plyne, že chyba při výpočtu E_2 je (-2)krát větší než chyba při výpočtu E_1, dále chyba při výpočtu E_3 je (-3)krát větší než chyba při výpočtu E_2 atd. To znamená, že chyba při výpočtu E_9 je $(-2)(-3)\dots(-9) = 9!$ větší než chyba v E_1. Tedy chyba v E_1, která je přibližně $4{,}412 \times 10^{-7}$, vede na chybu $9!\, 4{,}412 \times 10^{-7} \approx 0{,}1601$, což znamená dosti velkou chybu. Užitím uvedené rekurentní formule došlo ke značné kumulaci chyby.

Toto je obecný problém tzv. trojčlenných rekurentních formulí. Doporučený postup je následující. Přepišme uvedenou formuli ve tvaru

$$E_{n-1} = \frac{1 - E_n}{n}, \qquad n = \dots, 3, 2.$$

Pak chyba bude na každém kroku redukována faktorem $1/n$. Začneme větší hodnotou n a postupujeme zpětně. Je třeba ovšem „odhadnout“ tuto „startovací“ hodnotu. Všimněme si, že

$$E_n = \int_0^1 x^n \mathrm{e}^{x-1}\, \mathrm{d}x \le \int_0^1 x^n\, \mathrm{d}x = \frac{1}{n+1}.$$

Tedy pro $n = 20$ je $E_{20} \le 1/21$ a lze položit $E_{20} = 0$. Nyní užitím formule $E_{n-1} = (1 - E_n)/n$ dostaneme $E_9 = 0{,}0916123$ a tato hodnota má 6 platných cifer.

§ 0.5. Stabilita algoritmů

Z předchozích příkladů je zřejmé, že velká nepřesnost vypočtených výsledků byla způsobena užitím nevhodného algoritmu, neboť při změně algoritmu byly vypočtené výsledky zcela vyhovující. S tím souvisí otázky stability algoritmů.

Definice 0.2. Algoritmus se nazývá *stabilní*, jestliže vypočtené řešení je přesným řešením „blízkého“ problému, tj. řešením problému s blízkými vstupními daty.

Tento pojem stability vysvětlíme na následujícím příkladě.

Příklad 0.7. Víme, že $\mathrm{fl}(x + y) = (x + y)(1 + \delta) = x(1 + \delta) + y(1 + \delta) = x' + y'$. Tedy vypočtený součet dvou čísel x, y v pohyblivé řádové čárce je přesný součet jiných dvou čísel x' a y'. Jelikož $|\delta| \le \mu$, jsou čísla x' a y' blízká číslům x, y. Tedy operace sčítání dvou čísel v pohyblivé řádové čárce je stabilní.

O stabilitě konkrétních algoritmů pojednáme v dalších kapitolách při realizaci jednotlivých numerických metod.

Výše uvedenými příklady nechceme čtenáře odradit od studia a používání numerických metod. Cílem bylo pouze upozornit na možná „úskalí“ při realizaci numerických metod a ukázat způsoby, jak tyto problémy překonat. Velmi pěkně jsou některé patologické jevy v numerické matematice objasněny v monografii [15].

§ 0.6. Symbolika O, o

Závěrem této kapitoly ještě uvedeme symboliku O a o, která se často používá pro vyjádření chyb matematických výrazů (viz např. [9]).

Nechť φ je funkce (reálná nebo komplexní) definovaná v okolí bodu a (může být i ∞). Nechť ψ je funkce kladná v prstencovém okolí bodu a. Symbol

$$\varphi(x) = O(\psi(x)) \text{ pro } x \to a \tag{0.9}$$

značí, že

$$\limsup_{x\to a} \frac{|\varphi(x)|}{\psi(x)} < \infty\,.$$

Podobně symbol

$$\varphi(x) = o(\psi(x)) \text{ pro } x \to a \tag{0.10}$$

označuje, že

$$\lim_{x\to a} \frac{|\varphi(x)|}{\psi(x)} = 0\,.$$

Podobně je možné definovat výraz

$$a_n = O(b_n) \text{ nebo } a_n = o(b_n) \text{ pro } n \to \infty\,,$$

kde a_n, b_n jsou prvky posloupností.

Dodatek „pro $x \to a$“ se často vynechává, pokud je jasné, o které a se jedná. Je to zejména v případech $a = 0$ či $a = \infty$, případně u posloupností, kde je zřejmé, že $n \to \infty$. Často používaný výraz je také $O(h^k)$, resp. $o(h^k)$, kde $\psi(h) = h^k$, přičemž zpravidla $h \to 0$.

Při počítání s výrazy obsahující symboly O a o platí následující pravidla:

$$\begin{aligned}
O(\psi(x)) + O(\psi(x)) &= O(\psi(x)) \\
o(\psi(x)) + o(\psi(x)) &= o(\psi(x)) \\
O(\psi(x)) \cdot O(\vartheta(x)) &= O(\psi(x) \cdot \vartheta(x)) \\
O(\psi(x)) \cdot o(\vartheta(x)) &= o(\psi(x) \cdot \vartheta(x)) \\
o(\psi(x)) &= O(\psi(x))
\end{aligned}$$

Tyto rovnice nejsou symetrické, platí jen zleva doprava. Např. poslední rovnice značí, že funkce splňující rovnici (0.10) splňuje také rovnici (0.9). Opačně to ovšem neplatí.

Pokud za funkci $\psi(x)$ vezmeme konstantu 1, dostáváme výrazy $\varphi(x) = O(1)$ a $\varphi(x) = o(1)$. První z nich znamená, že funkce φ je omezená v okolí bodu a, druhý, že φ má limitu 0 v bodě a.

Cvičení k úvodní kapitole

1. Najděte primární chybu, která vznikne, jestliže přibližných čísel je použito k výpočtu:

   a) součtu n čísel

   b) součinu n čísel

   c) podílu dvou čísel

   d) mocniny čísla, kdy exponent je znám přesně

2. Nechť je dáno n čísel $\tilde{a}_1, \ldots, \tilde{a}_n$, kde $\tilde{a}_i$ je správně zaokrouhleno na d_i desetinných míst. Chceme spočítat součet na $d = \min_i d_i$ desetinných míst. Ukažte, že je výhodnější nejprve všechna čísla sečíst a výsledek zaokrouhlit na d míst, než napřed každé číslo $\tilde{a}_i$ zaokrouhlit na d míst a pak sečíst.

3. Buďte $\tilde{x}$ resp. $\tilde{y}$ čísla v absolutní hodnotě menší než 1 a správně zaokrouhlena na $2d$ resp. d desetinných míst. Nechť $|\tilde{x}| < |\tilde{y}|$.Chceme spočítat podíl $\tilde{x}/\tilde{y}$ na d desetinných míst. Ukažte, že použití $2d$-místného dělence je výhodnější, než když dělence nejdříve zaokrouhlíme na d míst a pak dělíme.

Kapitola 1

Normy vektorů a matic

Nechť $\mathbb{C}^n$ resp. $\mathbb{R}^n$ je vektorový prostor všech uspořádaných n-tic komplexních resp. reálných čísel. Prvky tohoto prostoru budeme zapisovat ve tvaru sloupcových vektorů.

Definice 1.1. Vektorová norma na $\mathbb{C}^n$ je funkce $\|\cdot\|$ (z $\mathbb{C}^n$ do $\mathbb{R}$) s následujícími vlastnostmi:

1) $\|\boldsymbol{x}\| \geq 0, \quad \forall \boldsymbol{x} \in \mathbb{C}^n$

2) $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{o}, \quad \boldsymbol{o} = (0, \ldots, 0)^T$

3) $\|\alpha \boldsymbol{x}\| = |\alpha| \, \|\boldsymbol{x}\|, \quad \forall \alpha \in \mathbb{C}, \quad \forall \boldsymbol{x} \in \mathbb{C}^n$

4) $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n.$

Příklady vektorových norem:

1) $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$ (eukleidovská norma)

2) $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$ (oktaedrická norma)

3) $\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$ (krychlová norma)

Každá vektorová norma indukuje metriku danou vztahem $\varrho(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$.

Připomeňme ještě definici konvergence posloupnosti vektorů vzhledem k dané normě.

Definice 1.2. Řekneme, že posloupnost $\{\boldsymbol{x}^k\}_{k=1}^{\infty}$ vektorů z $\mathbb{C}^n$ konverguje k vektoru $\boldsymbol{x} \in \mathbb{C}^n$ vzhledem k normě $\|\cdot\|$, jestliže pro libovolné $\varepsilon > 0$ existuje index $N = N(\varepsilon)$ tak, že

$$\|\boldsymbol{x}^k - \boldsymbol{x}\| < \varepsilon$$

pro $\forall\, k \geq N(\varepsilon)$.

Nechť A je čtvercová matice řádu n s reálnými resp. komplexními prvky, tj.

$$A = \begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ a_{21} & & & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix}.$$

Označme $\mathcal{M}_n$ třídu všech matic tohoto typu.

Matici A lze považovat za vektor dimenze n^2. Mohli bychom tedy definovat normu matice jako normu vektoru. Ale z hlediska pozdějších aplikací je vhodnější požadovat, aby norma matice splňovala další vlastnosti. Z těchto důvodů definujeme maticovou normu takto:

Definice 1.3. Maticová norma na množině $\mathcal{M}_n$ je reálná funkce $\|\cdot\|$ s těmito vlastnostmi:

1) $\|A\| \geq 0, \quad \forall A \in \mathcal{M}_n$

2) $\|A\| = 0 \Leftrightarrow A$ je nulová matice

3) $\|\alpha A\| = |\alpha|\,\|A\|, \quad \forall \alpha \in \mathbb{C}, \quad \forall A \in \mathcal{M}_n$

4) $\|A + B\| \leq \|A\| + \|B\|, \quad A, B \in \mathcal{M}_n$

5) $\|AB\| \leq \|A\|\,\|B\|, \quad A, B \in \mathcal{M}_n$

Vlastnost 5) se nazývá *multiplikativnost*.

Někdy je vhodné požadovat, aby norma matice nějakým způsobem „souvisela“ s normou vektoru. Tuto vlastnost nazýváme *souhlasnost* a její definice je následující:

Definice 1.4. Řekneme, že maticová norma $\|\cdot\|$ je *souhlasná* s danou vektorovou normou $\|\cdot\|_\varphi$, jestliže

$$\|A\boldsymbol{x}\|_\varphi \leq \|A\|\,\|\boldsymbol{x}\|_\varphi, \qquad \forall \boldsymbol{x} \in \mathbb{C}^n, \quad \forall A \in \mathcal{M}_n.$$

Věta 1.1. *Nechť $\|\cdot\|_\varphi$ je vektorová norma na $\mathbb{C}^n$. Pak číslo*

$$\|A\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi = 1} \|A\boldsymbol{x}\|_\varphi$$

je maticová norma souhlasná s danou vektorovou normou $\|\cdot\|_\varphi$.

Tato norma se nazývá *přidružená k dané vektorové normě.*

Důkaz. Norma je spojitá funkce vektoru $\boldsymbol{x}$. Protože $A\boldsymbol{x}$ je rovněž vektor, je funkce $\|A\boldsymbol{x}\|_\varphi$ spojitá, a tedy dosáhne na uzavřené omezené množině $\Omega = \{\boldsymbol{x} \colon \|\boldsymbol{x}\|_\varphi = 1\}$ svého maxima. To znamená, že existuje vektor $\boldsymbol{x}_0 \in \mathbb{C}^n$, $\|\boldsymbol{x}_0\|_\varphi = 1$, tak, že

$$\|A\boldsymbol{x}_0\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi = 1} \|A\boldsymbol{x}\|_\varphi.$$

Tím je dokázána existence čísla $\|A\|_\varphi$. Nyní ukážeme, že jsou splněny všechny axiomy maticové normy.

1) Nechť $A \neq O$, kde O je nulová matice. Pak existuje vektor $\hat{\boldsymbol{x}} \in \mathbb{C}^n$, $\|\hat{\boldsymbol{x}}\|_\varphi = 1$ takový, že $A\hat{\boldsymbol{x}} \neq \boldsymbol{o}$ a tedy $\|A\hat{\boldsymbol{x}}\|_\varphi > 0$. Proto

$$\|A\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi=1} \|A\boldsymbol{x}\|_\varphi \geq \|A\hat{\boldsymbol{x}}\|_\varphi > 0.$$

2) Je zřejmé, že $\|A\|_\varphi = 0 \Leftrightarrow A$ je nulová matice.

3) Pro libovolné $\alpha \in \mathbb{C}$ platí

$$\|\alpha A\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi=1} \|(\alpha A)\boldsymbol{x}\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi=1} |\alpha| \, \|A\boldsymbol{x}\|_\varphi = |\alpha| \, \|A\|_\varphi.$$

Než dokážeme, že je splněn čtvrtý axiom, ukážeme, že norma $\|A\|_\varphi$ je souhlasná s danou vektorovou normou.
Nechť $\boldsymbol{y} \neq \boldsymbol{o}$ je libovolný vektor z $\mathbb{C}^n$. Vektor $\boldsymbol{x} = \boldsymbol{y}/\|\boldsymbol{y}\|_\varphi$ má normu rovnu jedné. Pak

$$\|A\boldsymbol{y}\|_\varphi = \|A(\boldsymbol{x}\|\boldsymbol{y}\|_\varphi)\| = \|\boldsymbol{y}\|_\varphi \|A\boldsymbol{x}\|_\varphi \leq \|\boldsymbol{y}\|_\varphi \|A\|_\varphi.$$

Tedy norma $\|A\|_\varphi$ je souhlasná s danou vektorovou normou.

4) Nechť $\tilde{\boldsymbol{x}} \in \mathbb{C}^n$, $\|\tilde{\boldsymbol{x}}\|_\varphi = 1$, je takový vektor, že

$$\|(A+B)\tilde{\boldsymbol{x}}\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi=1} \|(A+B)\boldsymbol{x}\|_\varphi.$$

Je tedy

$$\begin{aligned} \|A+B\|_\varphi = \|(A+B)\tilde{\boldsymbol{x}}\|_\varphi &\leq \|A\tilde{\boldsymbol{x}}\|_\varphi + \|B\tilde{\boldsymbol{x}}\|_\varphi \leq \\ &\leq \max_{\|\boldsymbol{x}\|_\varphi=1} \|A\boldsymbol{x}\|_\varphi + \max_{\|\boldsymbol{x}\|_\varphi=1} \|B\boldsymbol{x}\|_\varphi = \|A\|_\varphi + \|B\|_\varphi. \end{aligned}$$

5) Pro matici AB najdeme vektor $\bar{\boldsymbol{x}} \in \mathbb{C}^n$, $\|\bar{\boldsymbol{x}}\|_\varphi = 1$, takový, že

$$\|(AB)\bar{\boldsymbol{x}}\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi=1} \|(AB)\boldsymbol{x}\|_\varphi.$$

Pak je

$$\begin{aligned} \|AB\|_\varphi = \|(AB)\bar{\boldsymbol{x}}\|_\varphi = \|A(B\bar{\boldsymbol{x}})\|_\varphi &\leq \|A\|_\varphi \|B\bar{\boldsymbol{x}}\|_\varphi \leq \\ &\leq \|A\|_\varphi \|B\|_\varphi \|\bar{\boldsymbol{x}}\|_\varphi = \|A\|_\varphi \|B\|_\varphi. \end{aligned}$$

□

Věta 1.2. *Přidružená maticová norma je nejvýše rovna libovolné maticové normě souhlasné s danou vektorovou normou.*

Důkaz. Nechť $\|\cdot\|$ je maticová norma souhlasná s danou vektorovou normou $\|\cdot\|_\varphi$. Pak platí

$$\|A\boldsymbol{x}\|_\varphi \leq \|A\| \, \|\boldsymbol{x}\|_\varphi.$$

Víme, že existuje takový vektor $\boldsymbol{x}_0 \in \mathbb{C}^n$, že $\|\boldsymbol{x}_0\|_\varphi = 1$, $\|A\|_\varphi = \|A\boldsymbol{x}_0\|_\varphi$ a odtud plyne

$$\|A\|_\varphi = \|A\boldsymbol{x}_0\|_\varphi \leq \|A\| \, \|\boldsymbol{x}_0\|_\varphi = \|A\|,$$

a tedy $\|A\|_\varphi \leq \|A\|$. □

Věta 1.3. *Nechť maticová norma $\|\cdot\|$ je souhlasná s danou vektorovou normou $\|\cdot\|_\varphi$. Pak pro všechna vlastní čísla λ matice A platí:*

$$|\lambda| \leq \|A\| \, .$$

Důkaz. Nechť $\boldsymbol{x}$ je vlastní vektor matice A odpovídající (nenulovému) vlastnímu číslu λ, tj. $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Pak je

$$\|\lambda\boldsymbol{x}\|_\varphi = |\lambda| \, \|\boldsymbol{x}\|_\varphi = \|A\boldsymbol{x}\|_\varphi \leq \|A\| \, \|\boldsymbol{x}\|_\varphi,$$

jelikož $\|\boldsymbol{x}\|_\varphi \neq 0$, je $|\lambda| \leq \|A\|$. □

Definice 1.5. Nechť $\lambda_1, \ldots, \lambda_n$ jsou vlastní čísla matice A. Číslo

$$\varrho(A) = \max_{1 \leq i \leq n} |\lambda_i|$$

se nazývá *spektrální poloměr* matice A.

Věta 1.4. *Nechť $A \in \mathcal{M}_n$. Přidružené maticové normy k vektorovým normám $\|\cdot\|_1$, $\|\cdot\|_\infty$, $\|\cdot\|_2$ jsou dány vztahy*

(i) $\|A\|_1 = \max\limits_{1 \leq j \leq n} \sum\limits_{i=1}^{n} |a_{ij}|$,

(ii) $\|A\|_\infty = \max\limits_{1 \leq i \leq n} \sum\limits_{j=1}^{n} |a_{ij}|$,

(iii) $\|A\|_2 = \sqrt{\varrho(A^*A)}$, *$\varrho(A^*A)$ je spektrální poloměr A^*A, kde $A^* = \overline{A}^T$, pro reálné matice je $A^* = A^T$.*

Důkaz viz [5].

Poznámka 1. Norma $\|A\|_2$ se nazývá *spektrální norma* matice A.

Nechť $E \in \mathcal{M}_n$ je jednotková matice. Zřejmě

$$\|E\|_\varphi = \max_{\|\boldsymbol{x}\|_\varphi = 1} \|E\boldsymbol{x}\|_\varphi = 1$$

a pro souhlasnou maticovou normu platí $\|E\| \geq 1$.

Důležitou normou souhlasnou s vektorovou normou $\|\cdot\|_2$ je Frobeniova norma:

$$\|A\|_F = \left(\sum_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}} .$$

Zřejmě platí $\|E\|_F = \sqrt{n}$.

Stopa matice A ($tr\, A$) je definována jako součet diagonálních prvků, $tr\, A = \sum_{i=1}^{n} a_{ii}$.

Odtud plyne, že $\|A\|_F^2 = tr(A^* A)$. Dále lze ukázat, že pro maticové normy platí tyto ekvivalentní vztahy ([5]):

1. $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\|A\|_\infty$
2. $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2$
3. $\frac{1}{\sqrt{n}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1$.

Příklad 1.1. Vypočtěte normy $\|A\|_2$, $\|A\|_1$, $\|A\|_\infty$ a $\|A\|_F$ pro matici

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

Řešení. Je zřejmě $\|A\|_1 = 7$, $\|A\|_\infty = 6$. Dále

$$A^T A = \begin{pmatrix} 10 & 10 \\ 10 & 20 \end{pmatrix} \Rightarrow \lambda_{1,2} = \frac{3 \pm 10\sqrt{5}}{2} \Rightarrow$$

$$\Rightarrow \varrho(A^T A) \doteq 12.680340 \Rightarrow \|A\|_2 \doteq 3.5609465 .$$

Dále $\|A\|_F = (1 + 9 + 4 + 16)^{1/2} \doteq 5.4772256$.

Příklad 1.2. Nechť R je reálná ortogonální matice, tj. $R^T R = E$, $R^T = R^{-1}$. Vypočtěte $\|R\|_2$ a $\|AR\|_2$, $A \in \mathcal{M}_n$.

Řešení. Je $\|R\|_2^2 = \varrho(R^T R) = \varrho(E) = 1$.

Dále

$$\|AR\|_2^2 = \varrho((AR)^T AR) = \varrho(R^T A^T AR) = \varrho(R^{-1} A^T AR) .$$

Transformace $R^{-1} A^T AR$ je podobnostní transformace, která nemění vlastní čísla matice. Odtud plyne, že spektrální poloměr matice $R^{-1} A^T AR$ je roven spektrálnímu poloměru matice $A^T A$. To znamená, že

$$\|AR\|_2^2 = \varrho(R^{-1} A^T AR) = \varrho(A^T A) = \|A\|_2^2 .$$

Věta 1.5. *Nechť $\|B\| < 1$, $\|\cdot\|$ je souhlasná s danou vektorovou normou. Pak matice $E - B$ je regulární a platí*

$$\left\|(E - B)^{-1}\right\| \leq \frac{\|E\|}{1 - \|B\|}.$$

Důkaz. Vlastní čísla matice B jsou řešením charakteristické rovnice $\det(B - \lambda E) = 0$.

Odtud

$$\begin{aligned} 0 = \det(B - \lambda E) &= \det(B - E + E - \lambda E) = \\ &= \det(B - E - (\lambda - 1)E). \end{aligned}$$

Tedy matice $B - E$ má vlastní čísla $\lambda - 1$. Protože $\varrho(B) \leq \|B\| < 1$, jsou všechna tato vlastní čísla různá od nuly a tedy matice $B - E$ i $E - B$ jsou regulární.

Nyní

$$E = (E - B)(E - B)^{-1} = (E - B)^{-1} - B(E - B)^{-1}.$$

Odtud

$$\begin{aligned} \|E\| &\geq \left\|(E - B)^{-1}\right\| - \left\|B(E - B)^{-1}\right\| \geq \\ &\geq \left\|(E - B)^{-1}\right\| - \|B\| \left\|(E - B)^{-1}\right\| \end{aligned}$$

a tedy

$$\left\|(E - B)^{-1}\right\| \leq \frac{\|E\|}{1 - \|B\|}.$$

□

Poznámka 2. Jestliže maticová norma uvažovaná v předchozí větě je přidruženou maticovou normou, pak

$$\left\|(E - B)^{-1}\right\| \leq \frac{1}{1 - \|B\|}.$$

Důsledek. *Je-li $\varrho(B) < 1$, je matice $E - B$ regulární.*

Důkaz plyne ihned z výše dokázaného faktu, že vlastní čísla matice $E - B$ jsou různá od nuly.

Cvičení ke kapitole 1

1. Nechť v rovině jsou dány dva vektory $\boldsymbol{u}$, $\boldsymbol{v}$. Najděte geometrické místo vektorů $\boldsymbol{w}$ takových, že
$$\|\boldsymbol{u} - \boldsymbol{w}\| = \|\boldsymbol{v} - \boldsymbol{w}\|.$$
Sestrojte tato geometrická místa pro normy $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ v případě, že $\boldsymbol{u} = (0,0)^T$, a) $\boldsymbol{v} = (1,1)^T$, b) $\boldsymbol{v} = (1, \frac{1}{2})^T$, c) $\boldsymbol{v} = (1,0)^T$.

2. Ověřte, že funkce $\|\cdot\|$ definovaná na množině $\mathcal{M}_n$ vztahem
$$\|A\| = \left(\sum_{i,j=1}^{n} |a_{ij}|^p \right)^{\frac{1}{p}}$$
je maticová norma právě tehdy, když $1 < p \leq 2$.

3. Nechť $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, kde $\boldsymbol{a}_j$ je j-tý sloupec matice A. Dokažte, že

$$\|A\|_F^2 = \sum_{i=1}^{n} \|\boldsymbol{a}_i\|_2^2 .$$

4. Pro matici $A = \begin{pmatrix} 2 & 1 & 1 \\ 2 & 3 & 2 \\ 1 & 1 & 2 \end{pmatrix}$ vypočtěte $\|A\|_1$, $\|A\|_2$ a $\|A\|_\infty$.

   (Řešení: $\|A\|_1 = 5$, $\|A\|_2 = \sqrt{\varrho(A^T A)} = 5,203\,527$, $\|A\|_\infty = 7$.)

5. Nechť k matici A existuje inverzní matice A^{-1}. Ukažte:

   a) Je-li $\lambda \neq 0$ vlastní číslo matice A, $\boldsymbol{x}$ je příslušný vlastní vektor, pak $1/\lambda$ je vlastní číslo matice A^{-1} s vlastním vektorem $\boldsymbol{x}$.

   b) Pro libovolnou přidruženou maticovou normu platí

$$\frac{1}{\|A^{-1}\|} \leq |\lambda|.$$

6. Nechť R je reálná ortogonální matice. dokažte, že pak $\|AR\|_F = \|A\|_F$.

   (*Řešení*: $\|AR\|_F^2 = tr((AR)^T AR) = tr(R^T A^T AR)$. Na druhé straně $tr(A) = \sum_{i=1}^{n} \lambda_i$, kde λ_i jsou vlastní čísla matice A. Jelikož matice $A^T A$ a $R^T A^T AR$ mají stejná vlastní čísla, je $tr(A^T A) = tr(R^T A^T AR)$ a tedy $\|AR\|_F = \|A\|_F$.)

7. Ukažte, že spektrální norma symetrické matice je rovna jejímu spektrálnímu poloměru.

   (*Řešení*: Je-li A symetrická matice, pak jsou všechna její vlastní čísla reálná. Jsou-li $\lambda_1,\ldots,\lambda_n$ vlastní čísla této matice, pak matice $A^2 = A^T A$ má vlastní čísla $\lambda_1^2,\ldots,\lambda_n^2$. Pak $\|A\|_2^2 = \varrho(A^T A) = \varrho(A^2) = \max\limits_{1\leq i\leq n} \lambda_i^2 = \varrho^2(A)$)

8. Nechť P a Q jsou ortogonální matice. Pak platí:

   a) $\|QAP\|_F = \|A\|_F$

   b) $\|QAP\|_2 = \|A\|_2$

   (*Řešení*:

   a) $\|QAP\|_F^2 = tr((QAP)^T QAP) = tr((AP)^T Q^T QAP) = tr((AP)^T AP)$, neboť $Q^T Q = E$.
   Stopa matice je invariantní vzhledem k podobnostní transformaci a z toho plyne (viz cvičení 6), že $tr((AP)^T AP) = tr(A^T A) = \|A\|_F^2$.

   b) $\|QAP\|_2^2 = \varrho((QAP)^T QAP) = \varrho((AP)^T Q^T QAP) = \varrho(P^T A^T AP) = \varrho(A^T A) = \|A\|_2^2$ (viz př. 1.2))

Kontrolní otázky ke kapitole 1

1. Může být maticová norma definována vztahem

$$\|A\| = \max_{1\le i,j\le n} |a_{ij}| \,?$$

   Ilustrujte na příkladě.

2. Je Frobeniova norma přidružená k vektorové normě $\|\cdot\|_2$?

3. Nechť $Q \in \mathcal{M}_n$ je ortogonální matice. Pak $\|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$. Dokažte. Platí toto tvrzení i pro normy $\|\cdot\|_1$, $\|\cdot\|_\infty$, $\|\cdot\|_F$?

Kapitola 2

Řešení nelineárních rovnic

Tato kapitola se bude zabývat numerickými metodami řešení nelineárních algebraických a transcendentních rovnic v případech, kdy přesné řešení nelze získat algebraickými metodami. Budeme se tedy zabývat hledáním kořenů, zejména reálných, rovnice

$$f(x) = 0, \tag{2.1}$$

kde x je reálná proměnná a f je v nějakém smyslu „rozumná“ funkce.

Číslo ξ, které je řešením rovnice (2.1) budeme nazývat *kořenem* funkce.

Při hledání kořenů lze postupovat takto:

A) *Separace kořenů*, tj. nalezení intervalů, ve kterých leží vždy právě jeden kořen rovnice (2.1).

B) *Zpřesnění* těchto kořenů.

Pro separaci kořenů lze užít známé věty z matematické analýzy:

Věta 2.1. *Nechť $f \in C[a,b]$ a nechť f nabývá v koncových bodech intervalu hodnot s opačnými znaménky, tj. $f(a)f(b) < 0$. Pak uvnitř tohoto intervalu leží alespoň jeden kořen rovnice* (2.1). *Jestliže existuje f' a má konstantní znaménko v tomto intervalu, pak existuje právě jeden kořen $\xi \in (a,b)$.*

Při separaci kořenů postupujeme tak, že nejdříve určíme znaménka funkce f v hraničních bodech jejího definičního oboru. Pak určujeme znaménka funkce v bodech, jejichž volba je určena chováním funkce f.

§ 2.1. Metoda bisekce

Na větě 2.1 je založena velmi jednoduchá numerická metoda pro nalezení kořenů — metoda *bisekce* neboli *metoda půlení*. Popišme nyní stručně tuto metodu.

Nechť $f \in C[a,b]$ a nechť $f(a)f(b) < 0$. Podle věty 2.1 leží v intervalu $[a,b]$ alespoň jeden kořen rovnice $f(x) = 0$. Předpokládejme pro jednoduchost, že tento kořen je jediný. Položme $a_0 = a$, $b_0 = b$, $s_0 = \frac{1}{2}(a_0 + b_0)$.

Je-li $f(a_0)f(s_0) < 0$, leží kořen v intervalu $[a_0, s_0]$ a položíme $a_1 = a_0$, $b_1 = s_0$ a postup opakujme pro interval $[a_1, b_1]$.

Je-li $f(s_0)f(b_0) < 0$, leží kořen v intervalu $[s_0, b_0]$ a položíme $s_0 = a_1$, $b_0 = b_1$ a postup opakujme pro interval $[a_1, b_1]$.

Je-li $f(s_0) = 0$, je $s_0 = \xi$ a kořen je nalezen.

Tímto způsobem dostaneme posloupnost intervalů

$$[a_0, b_0] \supset [a_1, b_1] \supset \ldots \supset [a_n, b_n] \supset \ldots,$$

přičemž $f(a_n)f(b_n) < 0, n = 0, 1, \ldots$

Pro koncové body těchto intervalů platí

$$a_0 \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq a_{n+1} \leq \ldots \leq \xi$$

$$\xi \leq \ldots \leq b_{n+1} \leq b_n \leq \ldots \leq b_0$$

a délky těchto intervalů jsou dány vztahem

$$b_n - a_n = \frac{b_0 - a_0}{2^n}, \qquad n = 1, 2, \ldots$$

Protože posloupnosti $\{a_n\}$, $\{b_n\}$ jsou omezené, monotonní a délka intervalů $[a_n, b_n]$ konverguje k nule, platí

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} a_n = \xi.$$

Nyní snadno ukážeme, že ξ je kořenem rovnice $f(x) = 0$. Funkce f je spojitá a platí $f(a_n)f(b_n) < 0, n = 0, 1, \ldots$. Odtud

$$\lim_{n\to\infty} f(b_n)f(a_n) = f^2(\xi) \leq 0,$$

ale odtud plyne, že $f(\xi) = 0$.

Z uvedeného postupu rovněž plyne, že

$$|s_n - \xi| \leq \frac{b-a}{2^{n+1}}, \qquad s_n = \frac{a_n + b_n}{2}.$$

Je totiž $\xi \in [a_n, b_n]$, $b_n - a_n = (b-a)/2^n$ a tudíž $|s_n - \xi| \leq (b_n - a_n)/2 = (b-a)/2^{n+1}$.

Uvedené úvahy můžeme zformulovat v následující větě.

Věta 2.2. *Nechť $f \in C[a,b]$, $f(a)f(b) < 0$ a nechť f má v intervalu $[a,b]$ jediný kořen ξ. Pak metoda bisekce generuje posloupnost $s_n = (a_n + b_n)/2$, $n = 0, 1, 2, \ldots$, která konverguje ke kořenu ξ a aproximuje kořen ξ takto:*

$$|s_n - \xi| \leq \frac{b-a}{2^{n+1}}. \tag{2.2}$$

Než uvedeme příklad na ilustraci metody bisekce, zmíníme se o problému zastavení výpočtu při použití numerické metody pro nalezení kořene. Předpokládejme, že numerická metoda generuje posloupnost $\{x^k\}$ konvergující ke kořenu ξ a nechť je dána požadovaná přesnost $\varepsilon > 0$. Jako kritérium pro zastavení výpočtu lze především doporučit

$$\left|\frac{x^{k+1} - x^k}{x^k}\right| < \varepsilon \tag{2.3}$$

nebo

$$|x^{k+1} - x^k| < \varepsilon \tag{2.4}$$

$$|f(x^k)| < \varepsilon \tag{2.5}$$

Kritéria (2.4) a (2.5) nejsou obecně vždy vhodná, neboť i když $|x^{k+1} - x^k| < \varepsilon$, nemusí také být $|x^{k+1} - \xi| < \varepsilon$ a totéž platí pro kritérium (2.5). Z těchto důvodů je nejvhodnějším kritériem pro zastavení výpočtů kritérium (2.3). Na druhé straně, u některých dále uvedených metod lze bez problémů použít kritérium (2.4), nebo je vhodné současné použití (2.4) a (2.5) (viz obrázek 2.1).

Rovnice $f(x) = x^3 - x - 1$ má podle věty 2.1 v intervalu $[1, 2]$ právě jeden kořen. Podle Cardanových vzorců je tento kořen ξ dán vztahem

$$\xi = \alpha + \beta, \qquad \alpha = \sqrt[3]{\frac{1}{2} + \frac{\sqrt{23}}{6\sqrt{3}}}, \qquad \beta = \sqrt[3]{\frac{1}{2} - \frac{\sqrt{23}}{6\sqrt{3}}},$$

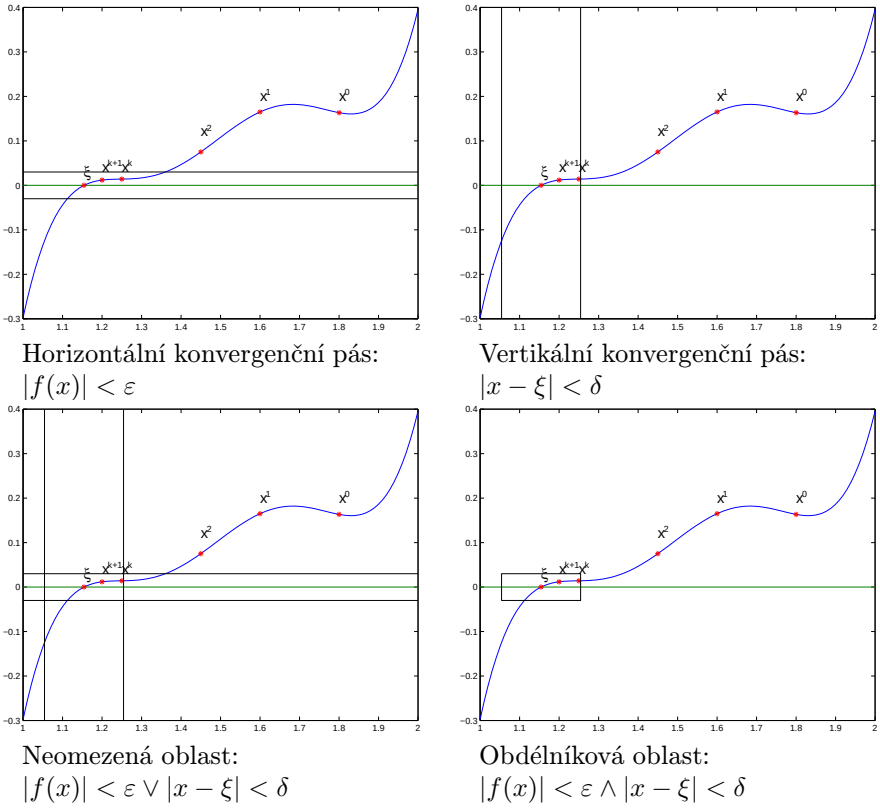
tj. $\xi \approx 1{,}3247179572447$.

Tuto rovnici budeme v této kapitole považovat v jistém smyslu za „testovací" pro jednotlivé metody, to znamená, že tyto metody budeme aplikovat na nalezení kořene této rovnice.

Příklad 2.1. Metodou bisekce najděte kořen funkce $f(x) = x^3 - x - 1$ ležící v intervalu $[1, 2]$ (obr. 2.2).

n	a_n	b_n	$b_n - a_n$
0	1,000000	2,000000	1,000000
1	1,000000	1,500000	0,500000
2	1,250000	1,500000	0,250000
3	1,250000	1,375000	0,125000
4	1,312500	1,375000	0,062500
5	1,312500	1,343750	0,031250
6	1,312500	1,328125	0,015625
7	1,320312	1,328125	0,007812

Je tedy $s_7 = 1{,}3242185$ a pro chybu aproximace platí $|s_7 - \xi| \leq 1/2^8$.



Obr. 2.1: Kritéria k zastavení iteračního procesu

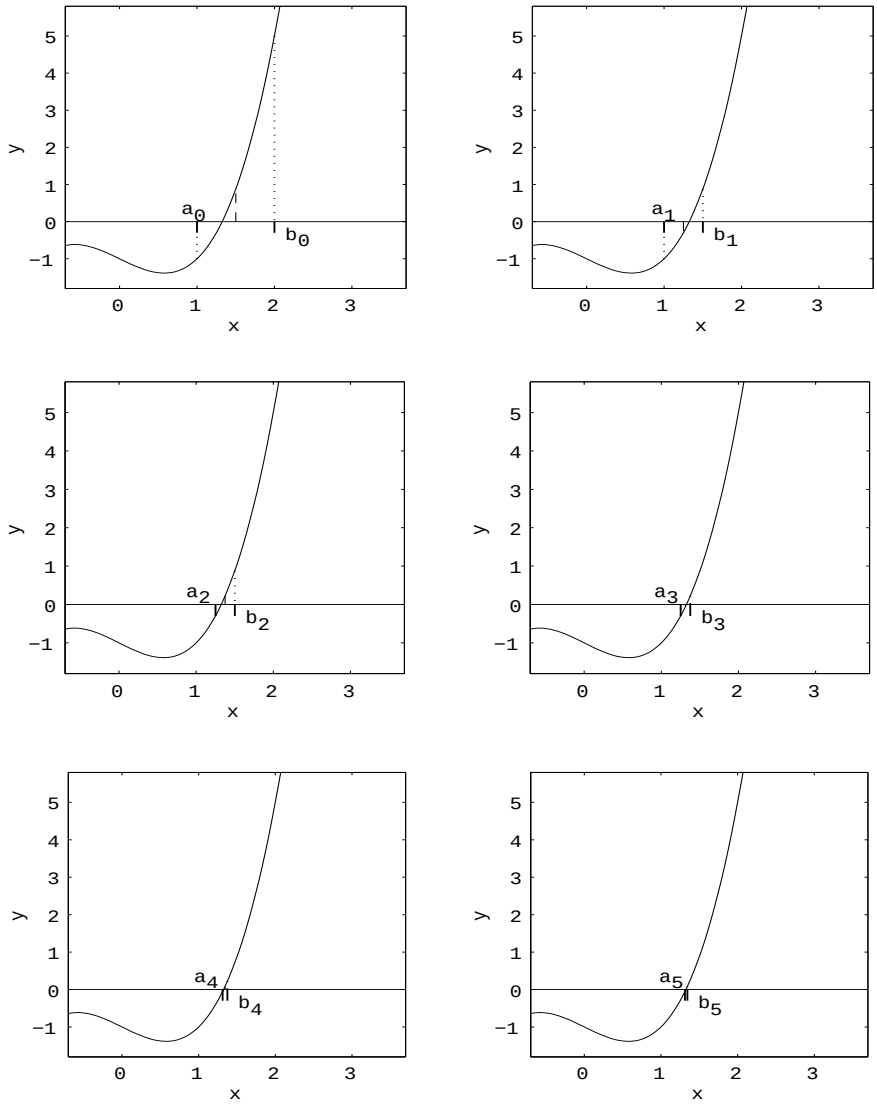
Poznámka 1. Při aplikaci metody bisekce je třeba věnovat pozornost ověření předpokladů: $f \in C[a,b]$, $f(a)f(b) < 0$. Následující příklad ukazuje, jaké problémy mohou nastat při nesplnění některého z předpokladů.

Příklad 2.2. Užitím metody bisekce najděte kořen funkce $f(x) = \frac{4x-7}{(x-2)^2}$.

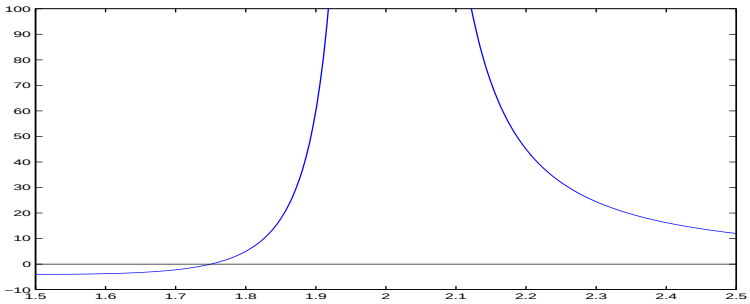
Řešení: Kořen ξ leží v intervalu $[1,5; 2,5]$, neboť $f(1,5) < 0$, $f(2,5) > 0$. Položme $a_0 = 1,5$, $b_0 = 2,5$. Pak $s_0 = (1,5 + 2,5)/2 = 2$, ale funkce $f(x) = (4x - 7)/(x-2)^2$ není definována v bodě $x = 2$. Metoda bisekce „selhala“, neboť daná funkce není spojitá na $[1,5; 2,5]$ (obr. 2.3). Vhodný interval pro použití metody bisekce je $[1,5; 1,9]$. Opět platí $f(1,5) < 0$, $f(1,9) > 0$, ale funkce $f \in C[1,5; 1,9]$. Metodou bisekce s počátečními hodnotami $a_0 = 1,5$, $b_0 = 1,9$ získáme posloupnost $s_0 = 1,7$, $s_1 = 1,8$, $s_2 = 1,75$. Hodnota $s_2 = 1,75$ je hledaný kořen ξ.

§ 2.2. Metoda prosté iterace

Nyní se budeme zabývat iteračními metodami pro nalezení kořenů rovnice (2.1).



Obr. 2.2: Metoda bisekce



Obr. 2.3: Graf funkce $f(x) = \frac{4x-7}{(x-2)^2}$

Tyto metody jsou založeny na řešení ekvivalentní úlohy $x = g(x)$, tj. na nalezení pevných bodů funkce g. Bod ξ je pevným bodem funkce g jestliže $g(\xi) = \xi$. Ekvivalentnost úloh $x = g(x)$ a $f(x) = 0$ znamená: jestliže ξ je *pevný bod funkce* g, pak ξ je kořen funkce f a naopak. Nejdříve se budeme zabývat iteračními metodami pro nalezení pevného bodu ξ a pak volbou vhodné funkce g.

Věta 2.3. *Nechť* $g \in C[a,b], g : [a,b] \to [a,b]$. *Pak funkce* g *má v intervalu* $[a,b]$ *pevný bod. Jestliže* g *splňuje navíc Lipschitzovu podmínku s konstantou* q, $0 \leq q < 1$

$$|g(x) - g(y)| \leq q\,|x - y|\,, \qquad \forall x, y \in [a,b],$$

pak g *má v intervalu jediný pevný bod.*

Důkaz. Jestliže $g(a) = a$ nebo $g(b) = b$, je existence pevného bodu zřejmá. Předpokládejme nyní, že $g(a) > a$, $g(b) < b$ a uvažujme funkci $h, h(x) = g(x) - x$. Zřejmě $h \in C[a,b]$ a dále

$$h(a) = g(a) - a > 0, \qquad h(b) = g(b) - b < 0.$$

Z vlastností spojitých funkcí plyne, že existuje bod $\xi \in (a,b)$ tak, že $h(\xi) = 0$, tj. $g(\xi) - \xi = 0 \Rightarrow \xi = g(\xi)$ a tedy ξ je pevný bod funkce g.

Nechť funkce g splňuje Lipschitzovu podmínku s konstantou q, $0 \leq q < 1$. Předpokládejme, že existují dva pevné body ξ, η. Nyní pro tyto body platí

$$|\xi - \eta| = |g(\xi) - g(\eta)| \leq q\,|\xi - \eta| < |\xi - \eta|\,,$$

což je spor a odtud plyne, že $\xi = \eta$. □

Důsledek. *Nechť* $g \in C^1[a,b]$, g: $[a,b] \to [a,b]$ *a*

$$|g'(x)| \leq q < 1, \qquad \forall x \in [a,b].$$

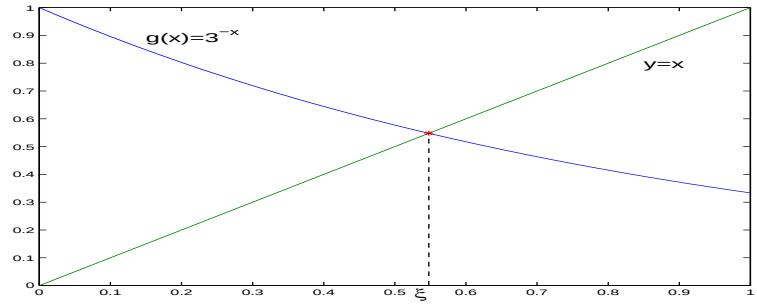
Pak g *má v intervalu* $[a,b]$ *jediný pevný bod.*

Důkaz ihned plyne aplikací věty o střední hodnotě:

$$|g(x) - g(y)| = |g'(\alpha)|\,|x - y| \leq q|x - y|, \qquad \alpha \in (a, b).$$

Poznámka 2. Předpoklady uvedené ve větě 2.3 jsou postačující, ale nikoliv nutné pro jednoznačnost pevného bodu.

Příklad 2.3. Je dána funkce $g(x) = 3^{-x}$, $g'(x) = -3^{-x} \ln 3 < 0$ na $[0, 1]$. Funkce g je tedy klesající na intervalu $[0, 1]$. Dále $g(1) = \frac{1}{3} \leq g(x) \leq 1 = g(0)$. Odtud plyne, že funkce g zobrazuje interval $[0, 1]$ do sebe. Dále $g'(0) = -\ln 3 \doteq -1,09861$, a tedy $|g'(x)| \not\leq q < 1$ na intervalu $[0, 1]$. Ale je jasné, že pevný bod je jediný, neboť g je klesající (viz obrázek). Jak je třeba „zúžit“ interval, aby byla splněna podmínka $|g'(x)| \leq q < 1$ (viz obrázek 2.4)?



Obr. 2.4: Graf funkce $g(x) = 3^{-x}$

Poznámka 3. *Řešit rovnici $x = g(x)$ geometricky znamená hledat průsečík přímky $y = x$ s křivkou $y = g(x)$.*

Zabývejme se nyní numerickými metodami určení pevného bodu funkce g.

Nechť $g \in C[a, b]$, $g : [a, b] \to [a, b]$ a zvolme libovolnou počáteční aproximaci $x^0 \in [a, b]$. Generujme posloupnost $\{x^k\}_{k=0}^{\infty}$ takto:

$$x^{k+1} = g(x^k), \qquad k = 0, 1, 2, \ldots \tag{2.6}$$

Funkci g nazýváme *iterační funkcí* a metodu (2.6) *iterační metodou* nebo také *metodou prosté iterace*.

Iterační metoda (2.6) patří mezi *jednokrokové* iterační metody, neboť výpočet x^{k+1} závisí pouze na jedné předchozí aproximaci x^k. Obecně jsou funkcionální iterační metody tvaru

$$x^{k+1} = g(x^k, x^{k-1}, \ldots, x^{k-j+1}), j \geq 2. \tag{2.7}$$

Tyto metody nazýváme *j-krokovými* metodami.

Otázkou nyní je, za jakých předpokladů bude iterační posloupnost[1] (2.6) resp. (2.7) konvergovat a jak rychle bude tato posloupnost konvergovat k pevnému bodu ξ.

V další části této kapitoly bude mít značný význam řád iterační metody jako „míra“ rychlosti konvergence metody. Definujme nejdříve chybu k-té iterace vztahem

$$e_k = x^k - \xi.$$

Předpokládejme nyní, že metoda (2.7) je konvergentní:

$$\lim_{k\to\infty} x^k = \xi.$$

Existuje-li nyní reálné číslo $p \geq 1$ takové, že platí

$$\lim_{k\to\infty} \frac{\left|x^{k+1} - \xi\right|}{|x^k - \xi|^p} = \lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|^p} = C \not\equiv 0,$$

řekneme, že daná iterační metoda je *řádu* p pro bod ξ. Konstanta C se nazývá asymptotickou konstantou chyby a závisí na funkci g. Požadavek $C \not\equiv 0$ znamená, že $C \neq 0$ pro obecnou funkci g. Tento požadavek zaručuje jednoznačnost čísla p. Jestliže pro nějakou funkci g je konstanta C rovna nule, pak iterační metoda konverguje rychleji než obvykle.

Zabývejme se nyní jednokrokovými iteračními metodami. Následující věta ukazuje, že řád těchto metod je přirozené číslo.

Věta 2.4. *Nechť funkce g má v okolí bodu ξ derivace až do řádu $p \geq 1$ včetně. Iterační metoda $x^{k+1} = g(x^k), k = 0, 1, \ldots$ je řádu p tehdy a jen tehdy, když platí*

$$\xi = g(\xi), \qquad g^{(j)}(\xi) = 0, \qquad 1 \leq j < p, \qquad g^{(p)}(\xi) \neq 0.$$

Důkaz. Vyjádříme funkci g v okolí bodu ξ pomocí Taylorova vzorce

$$g(x^k) = \xi + (x^k - \xi)g'(\xi) + \ldots + \frac{(x^k - \xi)^{p-1}}{(p-1)!} g^{(p-1)}(\xi) + \frac{(x^k - \xi)^p}{p!} g^{(p)}(\alpha) =$$
$$= \xi + \frac{(x^k - \xi)^p}{p!} g^{(p)}(\alpha), \tag{2.8}$$

kde bod α leží v intervalu určeném body x^k a ξ. Protože $x^{k+1} = g(x^k)$, dostaneme z předchozího vztahu

$$x^{k+1} - \xi = \frac{(x^k - \xi)^p}{p!} g^{(p)}(\alpha), \tag{2.9}$$

a tedy

$$\lim_{k\to\infty} \frac{\left|x^{k+1} - \xi\right|}{|x^k - \xi|^p} = \frac{\left|g^{(p)}(\xi)\right|}{p!} \not\equiv 0.$$

[1] Někdy také říkáme, že „iterační metoda konverguje“ místo „posloupnost konverguje“.

Metoda je tedy řádu $p \geq 1$, p přirozené číslo.

Z druhé strany: Nechť pro některé $j, 1 \leq j < p$, platí $g^{(j)}(\xi) \neq 0$. Pak z (2.8) plyne, že metoda nemůže být řádu p. Rovněž, jestliže $g^{(p)}(\xi) = 0$, pak z (2.9) plyne, že metoda není řádu p. □

Věta 2.5. *Nechť jsou splněny předpoklady věty* 2.3. *Pak pro libovolnou počáteční aproximaci* $x^0 \in [a,b]$ *je posloupnost* $\{x^k\}_{k=0}^{\infty}$, $x^k = g(x^{k-1})$, *konvergentní a platí* $\lim_{k\to\infty} x^k = \xi$, *kde* ξ *je pevný bod funkce* g.

Důkaz. Funkce g zobrazuje interval $[a,b]$ do sebe. Odtud plyne, že posloupnost $\{x^k\}_{k=0}^{\infty}$ je definována pro všechna $k \geq 0$ a $x^k \in [a,b]$ pro všechna $k \geq 0$. Dále

$$\left|x^k - \xi\right| = \left|g(x^{k-1}) - g(\xi)\right| \leq q\left|x^{k-1} - \xi\right|.$$

Indukcí odtud plyne, že

$$\left|x^k - \xi\right| \leq q^k\left|x^0 - \xi\right|.$$

Jelikož $0 \leq q < 1$, je

$$\lim_{k\to\infty}\left|x^k - \xi\right| = 0,$$

a tedy posloupnost $\left\{x^k\right\}_{k=0}^{\infty}$ konverguje k pevnému bodu ξ. □

Důsledek. *Nechť funkce* g *splňuje předpoklady věty* 2.3. *Pak pro posloupnost* $\{x^k\}_{k=0}^{\infty}$, $x^0 \in [a,b]$, $x^k = g(x^{k-1})$, *platí*

$$\left|x^k - \xi\right| \leq \frac{q^k}{1-q}\left|x^0 - x^1\right|, \qquad \forall k \geq 1. \tag{2.10}$$

Důkaz. Z konstrukce iterační posloupnosti plyne:

$$\left|x^{k+1} - x^k\right| = \left|g(x^k) - g(x^{k-1})\right| \leq q\left|x^k - x^{k-1}\right| \leq \ldots \leq q^k\left|x^1 - x^0\right|.$$

Dále pro $m > k \geq 1$

$$\begin{aligned}\left|x^m - x^k\right| &\leq \left|x^m - x^{m-1}\right| + \left|x^{m-1} - x^{m-2}\right| + \ldots + \left|x^{k+1} - x^k\right| \leq \\ &\leq q^{m-1}\left|x^1 - x^0\right| + q^{m-2}\left|x^1 - x^0\right| + \ldots + q^k\left|x^1 - x^0\right| = \\ &= q^k\left(1 + q + \ldots + q^{m-k-1}\right)\left|x^1 - x^0\right|.\end{aligned}$$

Jelikož jsou splněny předpoklady věty 2.5 o konvergenci iteračního procesu, je $\lim_{m\to\infty} x^m = \xi$ a platí

$$\left|\xi - x^k\right| = \lim_{m\to\infty}\left|x^m - x^k\right| \leq q^k\left|x^1 - x^0\right|\sum_{i=0}^{\infty} q^i = \frac{q^k}{1-q}\left|x^1 - x^0\right|.$$

□

Je zřejmé, že věta 2.5 *je důsledkem známé Banachovy věty o pevném bodě.*

Poznámka 4. Rychlost konvergence závisí na faktoru $q^k/(1-q)$. Je-li q malé, rychlost je větší. Pro q blízké 1 je konvergence pomalá. Vztahu (2.10) lze užít jako kriteria pro zastavení výpočtu.

Podívejme se nyní na problematiku iteračních procesů a pevných bodů z geometrického hlediska. Uvedeme klasifikaci pevných bodů. Tato klasifikace je poměrně hrubá, ale pro naše účely je postačující.

Definice 2.1. Pevný bod ξ funkce $g \in C[a,b]$ se nazývá

a) *přitahující* (atraktivní) pevný bod, jestliže existuje takové okolí V tohoto bodu ξ, že pro každou počáteční aproximaci $x^0 \in V$ posloupnost iterací $\{x^k\}_{k=0}^{\infty}$ konverguje k bodu ξ.

b) *odpuzující* (repulzivní) pevný bod, jestliže existuje takové okolí U bodu ξ, že pro každou počáteční aproximaci $x^0 \in U, x^0 \neq \xi$, existuje takové k, že $x^k \notin U$.

Následující věta uvádí, kdy je pevný bod přitahující a kdy je odpuzující.

Věta 2.6. *Nechť $g \in C[a,b], g : [a,b] \to [a,b]$ a nechť ξ je pevný bod.*

a) *Jestliže pro všechna $x \neq \xi$ z nějakého okolí V bodu ξ platí*

$$\left|\frac{g(x)-g(\xi)}{x-\xi}\right| < 1, \tag{2.11}$$

pak ξ je přitahující pevný bod.

b) *Jestliže pro všechna $x \neq \xi$ z nějakého okolí U bodu ξ platí*

$$\left|\frac{g(x)-g(\xi)}{x-\xi}\right| > 1, \tag{2.12}$$

pak ξ je odpuzující pevný bod.

Důkaz viz [20].

Důsledek. *Nechť $g \in C[a,b], g : [a,b] \to [a,b]$ a nechť g má v bodě ξ derivaci.*

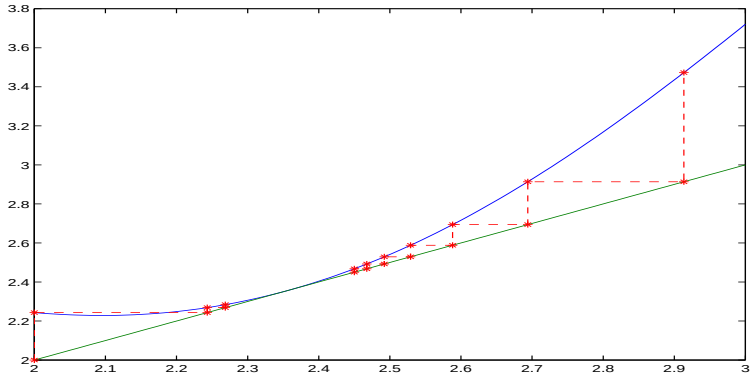
a) *Je-li $|g'(\xi)| < 1$, pak ξ je přitahující pevný bod.*

b) *Je-li $|g'(\xi)| > 1$, pak ξ je odpuzující pevný bod.*

Obrázky 2.8 a 2.9 znázorňují přitahující a odpuzující body funkce $g(x) = Ax(1-x)$. Případ $|g'(\xi)| = 1$ je třeba vyšetřovat zvlášť. Může nastat situace, že při počáteční iteraci na jedné straně okolí bodu ξ proces konverguje a na druhé straně diverguje (viz obr. 2.5).

Úmluva. Pro iterační proces platí

$$x^1 = g(x^0), \qquad x^2 = g(x^1) = g\left(g(x^0)\right), \qquad x^3 = g\left(g\left(g(x^0)\right)\right).$$



Obr. 2.5: Případ $g'(\xi) = 1$

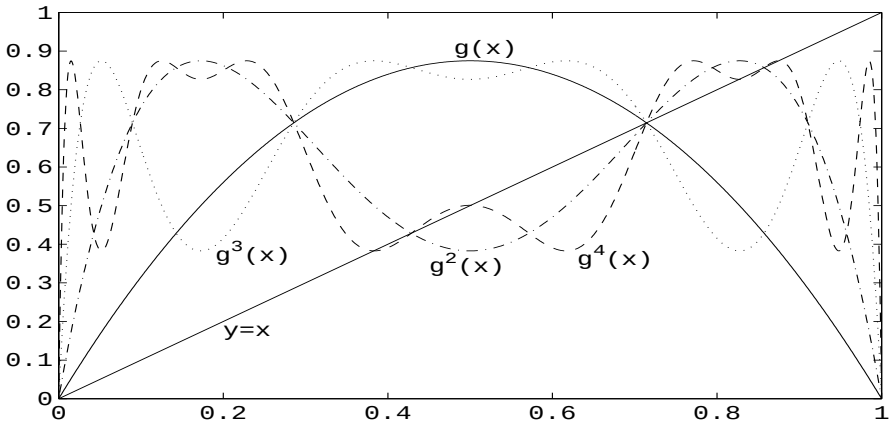
Obecně lze iteraci x^{k+1} definovat rekurzívně takto:

$$x^{k+1} = g^{k+1}(x^0),$$

přičemž

$$g^1(x) = g(x), \qquad g^{k+1}(x) = g\left(g^k(x)\right).$$

Funkce g^k se nazývá *k-tá iterace* funkce g (viz obr. 2.6).



Obr. 2.6: Grafy iterací funkce $g(x) = 3{,}5x(1-x)$

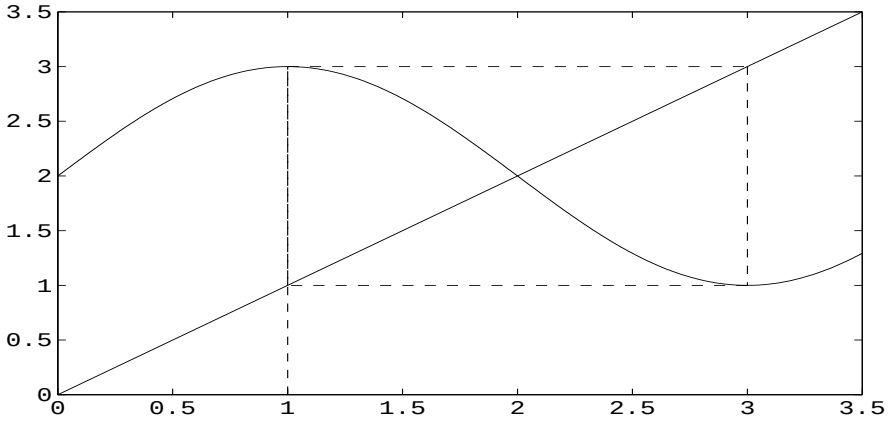
Doposud jsme se zabývali otázkami konvergence a divergence posloupnosti $\{x^k\}$. Ale někdy je užitečné zabývat se posloupnostmi, které jsou periodické. Základním pojmem je zde pojem cyklu a jeho řádu.

Definice 2.2. Nechť $g \in C[a,b], g\colon [a,b] \to [a,b]$. Řekneme, že bod $x^0 \in [a,b]$ je *bodem cyklu řádu n* funkce g, nebo že x^0 generuje *cyklus řádu n*, jestliže $g^n(x^0) = x^0$, $g^k(x^0) \neq x^0$ pro $k = 1, 2, \ldots, n-1$.

Poznámka 5. Je-li x^0 bod cyklu řádu n, pak je pevným bodem funkce g^n.

Uvažujme rovnici $x = \sin\frac{\pi}{2}x + 2$. Funkce $g(x) = \sin\frac{\pi}{2}x + 2$ má pevný bod $\xi = 2$. Zvolme počáteční aproximaci $x^0 = 1$. Pak $x^1 = g(x^0) = 3$, $x^2 = g^2(x^0) = 1$. Bod $x^0 = 1$ tedy generuje cyklus řádu 2.

Cyklus je ilustrován na obr. 2.7.



Obr. 2.7: Metoda prosté iterace, $x = \sin(\frac{\pi}{2}x) + 2$

Vyšetřujme nyní pevné body funkce $g(x) = Ax(1-x)$, $x \in [0,1]$, $A \in [0,4]$ v závislosti na parametru A. Tato funkce se používá na modelování některých biologických jevů. Za uvedených předpokladů je funkce g spojitá na intervalu $[0,1]$ a zobrazuje tento interval do sebe. Protože jsou splněny předpoklady první části věty 2.3, má funkce g v intervalu $[0,1]$ alespoň jeden pevný bod. V tomto jednoduchém případě pevné body snadno vypočteme a vyšetříme jejich vlastnosti.

Řešíme-li rovnici

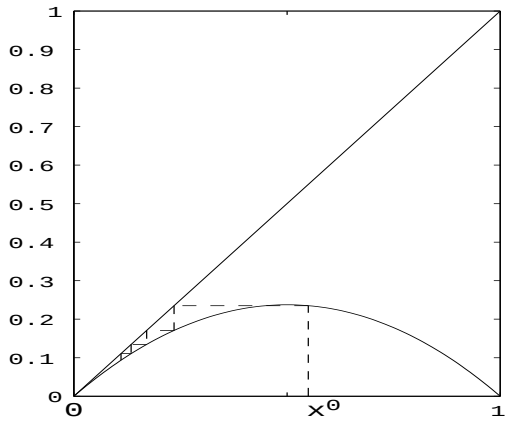
$$x = Ax(1-x)$$

dostaneme pro $A \in [0,1]$ právě jeden pevný bod $\xi_1 = 0$ a pro $A \in (1,4]$ právě dva pevné body $\xi_1 = 0$, $\xi_2 = 1 - 1/A$. Rozebereme nyní jednotlivé případy.

1. $A \in [0,1]$. Je $g'(x) = A - 2Ax$; $g'(0) = A \leq 1$. Funkce má jediný pevný bod $\xi_1 = 0$; graf funkce leží pod přímkou $y = x$ a bod $\xi_1 = 0$ je tedy přitahujícím pevným bodem (i pro $A = 1$) (viz obr. 2.8).

2. $A \in (1,3]$. V tomto případě $g'(\xi_1) = A > 1$; $g'(\xi_2) = A - 2$. To znamená, že bod ξ_1 je odpuzujícím pevným bodem a ξ_2 je pro $A \in (1,3)$ přitahujícím

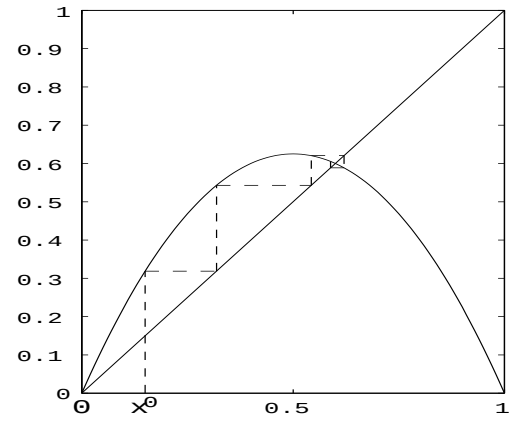
k	x^k	$g(x^k)$
0	0,5500	0,2351
1	0,2351	0,1708
2	0,1708	0,1346
3	0,1346	0,1106
4	0,1106	0,0935
5	0,0935	0,0805
6	0,0805	0,0703
7	0,0703	0,0621



Obr. 2.8: Metoda prosté iterace, $A = 0{,}95$, $\xi_1 = 0$

pevným bodem, neboť $|g'(\xi_2)| < 1$. Pro $A = 3$ je $|g'(\xi_2)| = 1$, ale i v tomto případě lze ukázat, že posloupnost $\{x^k\}$ bude konvergovat k pevnému bodu ξ_2, i když konvergence bude pomalá (viz [20] a obr. 2.9).

k	x^k	$g(x^k)$
0	0,1500	0,3187
1	0,3187	0,5429
2	0,5429	0,6204
3	0,6204	0,5888
4	0,5888	0,6053
5	0,6053	0,5973
6	0,5973	0,6013
7	0,6013	0,5993

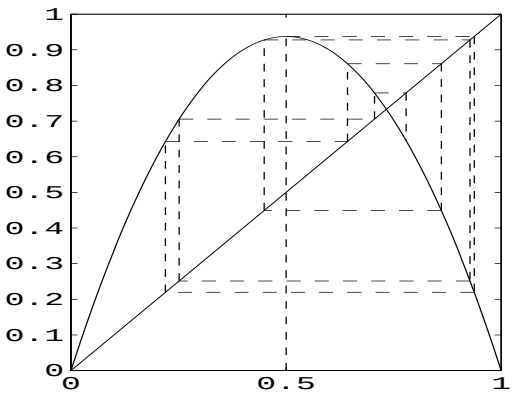


Obr. 2.9: Metoda prosté iterace, $A = 2{,}5$, $\xi_2 = 0{,}6$

3. $A \in (3, 4]$. Pro tyto hodnoty A již existují cykly různých řádů. Pro jisté hodnoty A existují posloupnosti generované libovolným prvkem $x^0 \in [0, 1]$, které jsou buď periodické nebo konvergují k periodické posloupnosti. Existuje také kritická hodnota $A = A_c = 3{,}5700\ldots$ taková, že pro $A > A_c$ lze vždy najít takovou počáteční aproximaci $x^0 \in [0, 1]$, že v odpovídající iterační posloupnosti neexistuje žádná zákonitost. Tato posloupnost může být dokonce tak neuspořádaná, že ji lze pokládat za posloupnost náhodných čísel a tomuto

jevu říkáme *chaos* (viz obr. 2.10). Zde se nebudeme podrobně zabývat těmito otázkami. Podrobnější informace lze najít např. v [20].

k	x^k	$g(x^k)$
0	0,500000	0,937500
1	0,937500	0,219727
2	0,219727	0,642926
3	0,642926	0,860896
4	0,860896	0,449077
5	0,449077	0,927776
6	0,927776	0,251279
7	0,251279	0,705518
8	0,705518	0,779109
9	0,779109	0,645367
10	0,645367	0,858256
11	0,858256	0,456197
12	0,456197	0,930305



Obr. 2.10: Metoda prosté iterace, $A = 3{,}75$, $\xi_2 = \frac{11}{15}$

Jako příklad ukážeme, že pro každé $A \in (3, 4]$ existuje alespoň jedna dvojice bodů x_1^0, x_2^0, které generují cyklus řádu 2.

Je zřejmé, že body cyklu řádu 2 najdeme řešením rovnice $x = g^2(x)$, tj.

$$A\ \ (Ax(1-x))\,(1-Ax(1-x)) = x. \tag{2.13}$$

Úpravou dostaneme

$$A^3x^4 - 2A^3x^3 + A^2(A+1)x^2 - A^2x + x = 0.$$

Pevné body $\xi_1 = 0$, $\xi_2 = 1 - 1/A$ jsou rovněž řešení této rovnice. Vydělíme tuto rovnici polynomem $(x-0)(x-(1-1/A))$ a výsledná rovnice je tvaru

$$A^2x^2 - A(A+1)x + A + 1 = 0. \tag{2.14}$$

Její diskriminant $D = A^2(A+1)(A-3)$. Odtud plyne:

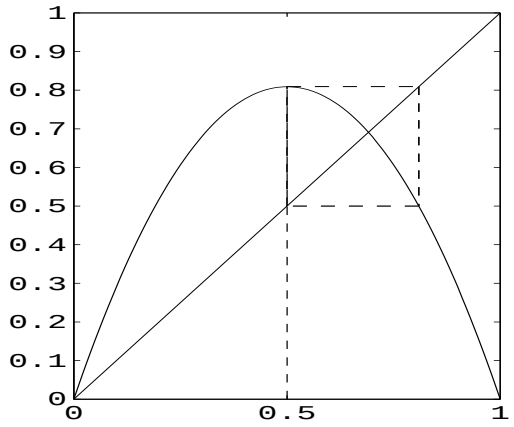
$0 \le A < 3 \quad \Rightarrow \quad D < 0 \quad \Rightarrow$ neexistuje bod generující cyklus řádu 2.

$A = 3 \quad \Rightarrow \quad D = 0 \quad \Rightarrow$ kvadratická rovnice má dvojnásobný kořen $\xi_2 = 1 - 1/A = 2/3$.

$3 < A \le 4 \quad \Rightarrow \quad D > 0 \quad \Rightarrow$ kvadratická rovnice má dva reálné různé kořeny x_1^0, x_2^0, které generují cyklus řádu 2.

Například lze snadno ověřit pro $A = 1 + \sqrt{5}$ jsou body cyklu řádu 2 body $x_1^0 = 0{,}5$; $x_2^0 = (3+\sqrt{5})/(2(1+\sqrt{5}))$ (viz obr. 2.11).

k	x^k	$g(x^k)$
0	0,5000	0,8090
1	0,8090	0,5000
2	0,5000	0,8090
3	0,8090	0,5000
4	0,5000	0,8090
5	0,8090	0,5000
6	0,5000	0,8090
7	0,8090	0,5000



Obr. 2.11: Metoda prosté iterace, $A = 1 + \sqrt{5}$, $\xi_2 = \sqrt{5}/(1 + \sqrt{5})$

Pro určitou hodnotu $A > 3$ vznikne první 4-cyklus. Od hodnoty $A = A_c = 3{,}5700\ldots$ se objevují cykly řádu $2^i p$, $p > 1$ je liché číslo. Pro hodnotu $A = 1 + \sqrt{8}$ vznikne první 3-cyklus.

Otázkami cyklů se obecně zabýval A. N. Šarkovskij. Uvedeme bez důkazu jeho známou větu (podrobněji viz [20]).

Věta 2.7. (Šarkovského věta). *Nechť $g \in C[a,b]$, $g\colon [a,b] \to [a,b]$. Na množině přirozených čísel definujme uspořádání takto:*

$$3 \prec 5 \prec 7 \prec \ldots \prec 2.3 \prec 2.5 \ldots \prec 2^i.3 \prec 2^i.5 \prec \ldots$$

$$\ldots \prec 2^{j+1} \prec 2^j \ldots \prec 8 \prec 4 \prec 2 \prec 1.$$

Jestliže g má cyklus řádu m, $m \prec n$, pak má g cyklus řádu n.

§ 2.3. Hledání vhodného tvaru iterační funkce

Zabývejme se nyní volbou vhodné iterační funkce g pro řešení rovnice $f(x) = 0$. Jednou z možností je „vhodně“ vypočítat x z rovnice $f(x) = 0$. Ale tento postup není vždy jednoduchý. Některé možné postupy ukazují následující příklady:

Příklad 2.4.

1. Najděte vhodnou iterační funkci pro nalezení největšího kladného kořene rovnice $x^3 + x - 1000 = 0$.
   Řešení. $g(x) = \sqrt[3]{1000 - x}$.

2. Pro rovnici $x - \operatorname{tg} x = 0$ najděte vhodnou iterační funkci pro určení nejmenšího kladného kořene.
   Řešení. Nejmenší kladný kořen leží v intervalu $[\pi, \frac{3}{2}\pi)$ a vhodné iterační funkce jsou například

(a) $g(x) = \dfrac{1}{\operatorname{tg} x} - \dfrac{1}{x} + x,$

(b) $g(x) = \operatorname{arctg} x + \pi.$

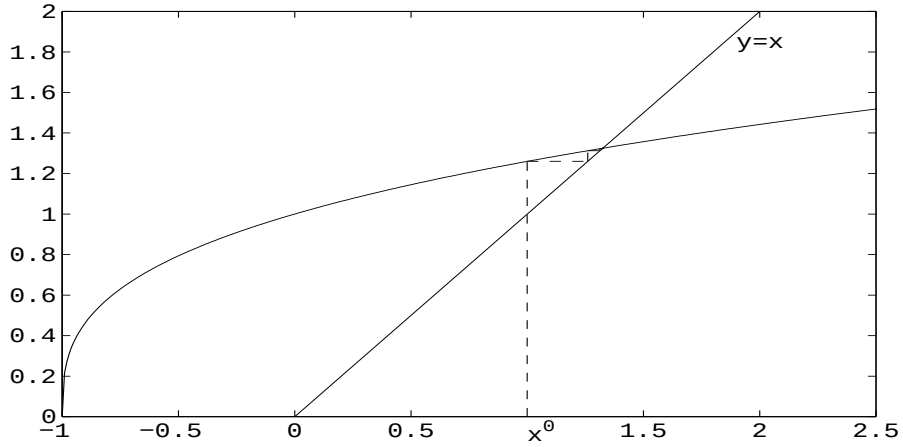
Příklad 2.5. Pro funkci z příkladu 2.1, tj. $f(x) = x^3 - x - 1$, najděte vhodnou iterační funkci pro kořen $\xi \in [1, 2]$. Obrázky 2.12 a 2.13 ilustrují chování iterační posloupnosti pro různé volby iterační funkce g. Je zřejmé, že vhodná iterační funkce je funkce $g(x) = (x+1)^{\frac{1}{3}}$.

a) $g(x) = (x+1)^{\frac{1}{3}}$

$x^0 = 1$
$x^1 = 1{,}259921050$
$x^2 = 1{,}312293837$
$x^3 = 1{,}322353819$
$x^4 = 1{,}324268745$
$x^5 = 1{,}324632625$
$x^6 = 1{,}324701749$
$x^7 = 1{,}324714878$
$x^8 = 1{,}324717372$
$x^9 = 1{,}324717846$

b) $g(x) = x^3 - 1$

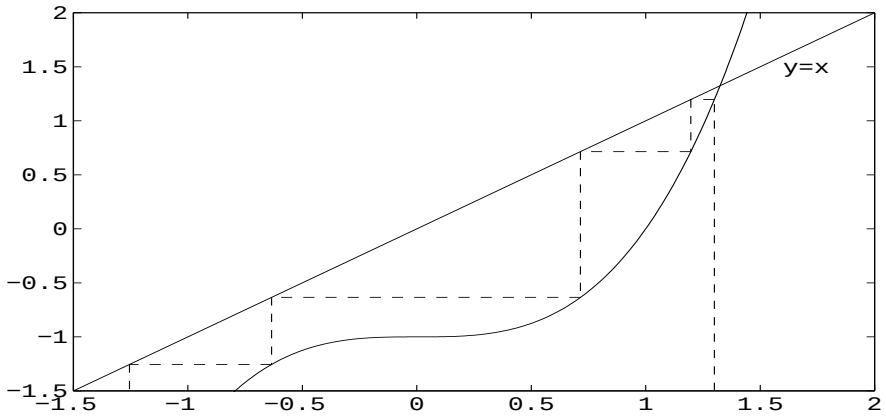
$x^0 = \ \ 1{,}3$
$x^1 = \ \ 1{,}197$
$x^2 = \ \ 0{,}715072373$
$x^3 = -0{,}634363117$
$x^4 = -1{,}255278226$
$x^5 = -2{,}977971306$
$x^6 = -27{,}40958194$
$x^7 = -20593{,}41275$



Obr. 2.12: Metoda prosté iterace, $x = (x+1)^{\frac{1}{3}}$, $x^0 = 1$

Podle věty 2.3 je třeba najít takovou funkci g, pro kterou $|g'(x)| \leq q < 1$ v okolí bodu ξ.

Lze snadno ověřit, že pro iterační funkci $g(x) = x^3 - 1$ nejsou splněny předpoklady věty 2.3 a pro iterační funkci $g(x) = (x+1)^{\frac{1}{3}}$ jsou tyto předpoklady splněny $(q = 1/3\sqrt[3]{4})$.



Obr. 2.13: Metoda prosté iterace, $x = x^3 - 1$, $x^0 = 1{,}3$

Příklad 2.6. Najděte vhodnou iterační funkci pro výpočet hodnoty $\sqrt[3]{25}$.
Řešení. Zvolme funkci $g(x) = \frac{5}{\sqrt{x}}$ za iterační funkci. Pevným bodem této funkce je bod $\xi = \sqrt[3]{25}$. Ukážeme, že funkce g splňuje na intervalu $I = [2,6; 3,5]$ předpoklady věty 2.3.
Je

$$g'(x) = -\frac{5}{2}x^{-\frac{3}{2}} \quad \Rightarrow \quad g \text{ je klesající na intervalu } I,$$

a jelikož $g(2,6) \doteq 3,101$, $g(3,5) \doteq 2,673$, zobrazuje g tento interval do sebe.
Dále

$$g''(x) = \frac{15}{4}x^{-\frac{5}{2}}, \qquad g'''(x) = -\frac{75}{8}x^{-\frac{7}{2}}.$$

Odtud plyne, že funkce g' je rostoucí a konkávní na I. Tedy

$$\max_{x \in I} |g'(x)| = |g'(2,6)| \doteq 0,596$$

Funkce g tedy splňuje předpoklady věty 2.3 s $q \doteq 0,596$ a z věty 2.5 plyne konvergence pro každou počáteční aproximaci $x^0 \in I$. Zvolme počáteční aproximaci $x^0 = 3$; další aproximace jsou generovány vztahem $x^{k+1} = 5/\sqrt{x^k}$:
$x^1 \doteq 2,886751$, $x^2 \doteq 2,942831$, $x^3 \doteq 2,914656$, $x^4 \doteq 2,928709$, $x^5 \doteq 2,921675$, atd.
Odhad relativní chyby pro x^4 je roven

$$\left| \frac{x^5 - x^4}{x^4} \right| \doteq 0,0024\,.$$

Z uvedených příkladů je vidět, že při hledání řešení rovnice $f(x) = 0$ je často možné vytvořit iterační funkci g více způsoby, přičemž ne vždy její vlastnosti

zaručují konvergenci iteračního procesu. Ve většině případů je ale možné volit iterační funkci ve tvaru

$$g(x) = x - Mf(x),$$

kde vhodnou volbou konstanty M lze často zaručit splnění předpokladů věty 2.3. Iterační proces má pak tvar

$$x^{k+1} = x^k - Mf(x^k), \quad M \neq 0.$$

Geometricky je bod x^{k+1} průsečík osy x a přímky procházející bodem $(x^k, f(x^k))$ se směrnicí $1/M$.

Příklad 2.7. Najděte vhodnou iterační funkci pro výpočet hodnoty $\ln 2$.
Řešení. Řešíme rovnici $e^x = 2$, tj. $f(x) = e^x - 2 = 0$. Dále víme, že $x \in I = [0, 1]$. Položme

$$g(x) = x - Mf(x) = x - M(e^x - 2)\,.$$

Nejprve zkoumejme derivaci funkce g:

$$g'(x) = 1 - Me^x, \quad |g'(x)| = |1 - Me^x|.$$

Pokud chceme, aby $|g'(x)| \leq q < 1$ na $[0, 1]$, musí platit $0 < Me^x < 2$ pro $x \in I$, tedy $0 < M < 2/e$.
Z rovnice $g'(x) = 0$ dále lehce zjistíme, že funkce g může mít lokální extrém jedině v bodě $\ln(1/M)$, pokud ale zvolíme $M < 1/e$, leží tento bod mimo interval I a g je na I monotonní. Tedy k tomu, aby se interval I zobrazil do sebe, stačí, když se do něj zobrazí krajní body:

$$g(0) = -M(1 - 2) = M \in [0, 1], \quad g(1) = 1 - M(e - 2) \in [0, 1].$$

Pro $M < 1/e$ oba vztahy platí, takže můžeme zvolit hodnotu M rovnu např. $1/3$.

§ 2.4. Newtonova metoda

Určit vhodnou iterační funkci může být obtížné. Z tohoto důvodu se budeme zabývat obecnými postupy, které rovnici $f(x) = 0$ „přiřadí“ za jistých předpokladů o funkci f vhodnou iterační funkci. Předpokládejme, že rovnice $f(x) = 0$ má jednoduchý kořen ξ, tj. $f'(\xi) \neq 0$. Pak pro funkci

$$g(x) = x - \frac{f(x)}{f'(x)} \tag{2.15}$$

je ξ pevným bodem ($\xi = g(\xi)$). Iterační metoda určená touto iterační funkcí je tvaru

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}, \tag{2.16}$$

(za předpokladu $f'(x^k) \neq 0$, $k = 0, 1, \ldots$) a nazývá se *Newtonova metoda*. Tato metoda má jednoduchý geometrický význam: *bod x^{k+1} je průsečík tečny ke grafu*

funkce f v bodě $[x^k, f(x^k)]$ s osou x. Z tohoto důvodu se Newtonova metoda také nazývá *metoda tečen*.

Věta 2.8. *Nechť $f \in C^2[a,b]$. Nechť $\xi \in [a,b]$ je kořenem rovnice $f(x) = 0$ a $f'(\xi) \neq 0$. Pak existuje $\delta > 0$ tak, že posloupnost $\{x^k\}_{k=0}^{\infty}$ generovaná Newtonovou metodou konverguje k bodu ξ pro každou počáteční aproximaci $x^0 \in [\xi - \delta, \xi + \delta] \subseteq [a,b]$.*

Důkaz. Ukážeme, že existuje subinterval $[\xi - \delta, \xi + \delta] \subseteq [a,b]$, na kterém iterační funkce (2.15) splňuje předpoklady věty 2.5.

Jelikož $f'(\xi) \neq 0$ a f' je spojitá na intervalu $[a,b]$, existuje takové $\delta_1 > 0$, že pro všechna $x \in [\xi - \delta_1, \xi + \delta_1] \subseteq [a,b]$ je $f'(x) \neq 0$. Tedy funkce g je definována a spojitá na intervalu $[\xi - \delta_1, \xi + \delta_1]$. Dále

$$g'(x) = 1 - \frac{f'^2(x) - f(x)f''(x)}{f'^2(x)} = \frac{f(x)f''(x)}{f'^2(x)}$$

pro $x \in [\xi - \delta_1, \xi + \delta_1]$, a protože $f \in C^2[a,b]$, je $g \in C^1[\xi - \delta_1, \xi + \delta_1]$. Podle předpokladu je $f(\xi) = 0$, a tedy

$$g'(\xi) = \frac{f(\xi)f''(\xi)}{f'^2(\xi)} = 0. \tag{2.17}$$

Funkce $g \in C^1[\xi - \delta_1, \xi + \delta_1]$, a tedy z (2.17) plyne, že existuje δ, $0 < \delta < \delta_1$, tak, že

$$|g'(x)| \leq q < 1$$

pro všechna $x \in [\xi - \delta, \xi + \delta]$.

Je třeba ještě ukázat, že $g\colon [\xi - \delta, \xi + \delta] \to [\xi - \delta, \xi + \delta]$, což plyne ihned aplikací věty o střední hodnotě, neboť pro libovolný bod $x \in [\xi - \delta, \xi + \delta]$ platí

$$|g(x) - \xi| = |g(x) - g(\xi)| = |g'(\alpha)(x - \xi)|,$$

a protože α leží v intervalu určeném body x a ξ, je $|g'(\alpha)| \leq q < 1$. Odtud $|g(x) - \xi| \leq q|x - \xi| < |x - \xi| \leq \delta$, a tedy $g(x) \in [\xi - \delta, \xi + \delta]$.

Funkce g splňuje na intervalu $[\xi - \delta, \xi + \delta]$ předpoklady věty 2.5 a to znamená, že posloupnost $\{x^k\}_{k=0}^{\infty}$ generovaná Newtonovou metodou konverguje pro každou počáteční aproximaci $x^0 \in [\xi - \delta, \xi + \delta]$ ke kořenu ξ. □

Důsledek. *Newtonova metoda je metoda druhého řádu pro jednoduchý kořen ξ.*

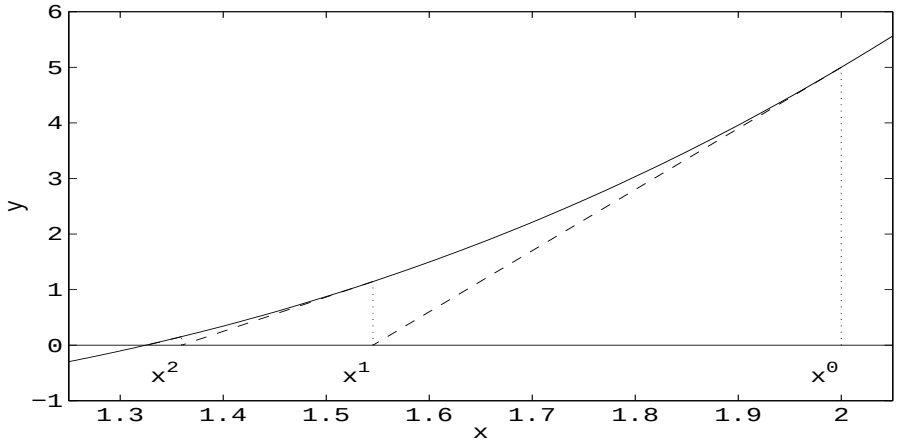
Důkaz plyne ihned ze vztahu (2.17), neboť $\xi = g(\xi)$, $g'(\xi) = 0$ a $g''(\xi) \neq 0$.

Příklad 2.8. Newtonovou metodou nalezněte kořen rovnice $f(x) = x^3 - x - 1$ ležící v intervalu $[1,2]$.

Řešení. Newtonova metoda je v tomto případě tvaru

$$x^{k+1} = x^k - \frac{(x^k)^3 - x^k - 1}{3(x^k)^2 - 1}.$$

k	x^k	$f(x^k)$
0	2	5
1	$1,\overline{54}$	1,145755071
2	1,359614916	0,153704934
3	1,325801345	0,004624917
4	1,324719049	0,000004658
5	1,324717957	$2,2204^{-16}$



Obr. 2.14: Newtonova metoda, $x^3 - x - 1 = 0$, $x^0 = 2$

Za počáteční aproximaci zvolme $x^0 = 2$. Metoda je ilustrována na obr. 2.14.

Jelikož $|x^5 - x^4|/x^5 = 8.10^{-7}$ a funkční hodnota $f(x^5) \approx 5.10^{-12}$, je x^5 velmi dobrou aproximací hledaného kořene ξ, což je na druhé straně vidět přímým porovnáním s hodnotou vypočtenou pomocí Cardanových vzorců (viz př. 2.1).

Předpoklady uvedené ve větě 2.8 znamenají, že Newtonova metoda bude konvergovat, zvolíme-li počáteční aproximaci x^0 „dostatečně“ blízko bodu ξ. V další části této kapitoly uvedeme metody, které jsou vhodné pro určení „dostatečně“ blízké počáteční aproximace. Jak jsme již viděli, jednou z takových metod je metoda bisekce.

Věta 2.9. *Nechť jsou splněny předpoklady předchozí věty. Pak pro posloupnost* $\{x^k\}_{k=0}^{\infty}$ *generovanou Newtonovou metodou platí*

$$\text{a)} \qquad |x^{k+1} - \xi| \le \frac{M}{2m}(x^k - \xi)^2, \tag{2.18}$$

$$\text{b)} \qquad |x^{k+1} - \xi| \le \frac{M}{2m}(x^{k+1} - x^k)^2, \tag{2.19}$$

kde $M = \max_{x \in I} |f''(x)|$, $m = \min_{x \in I} |f'(x)| > 0$, $I = [\xi - \delta, \xi + \delta]$.

Důkaz.

a) Užijme Taylorova vzorce ve tvaru

$$0 = f(\xi) = f(x^k) + f'(x^k)(\xi - x^k) + f''(\eta^k)\frac{(\xi - x^k)^2}{2},$$

kde bod η^k leží mezi body x^k, ξ.

Z Newtonovy metody plyne

$$x^{k+1} f'(x^k) = x^k f'(x^k) - f(x^k).$$

Po dosazení do předchozího vztahu dostaneme

$$0 = -x^{k+1} f'(x^k) + \xi f'(x^k) + f''(\eta^k)\frac{(\xi - x^k)^2}{2}$$

a odtud

$$\left|x^{k+1} - \xi\right| = \frac{\left|f''(\eta^k)\right|}{2|f'(x^k)|}(\xi - x^k)^2 \leq \frac{M}{2m}(\xi - x^k)^2.$$

b) Pro důkaz vztahu (2.19) použijeme opět Taylorova vzorce:

$$f(x^{k+1}) = f(x^k) + f'(x^k)(x^{k+1} - x^k) + \frac{f''(\alpha^k)}{2}(x^{k+1} - x^k)^2,$$

kde α^k leží mezi body x^k a x^{k+1}.

Opět z Newtonovy metody plyne, že

$$f(x^k) + f'(x^k)(x^{k+1} - x^k) = 0$$

a tudíž

$$f(x^{k+1}) = \frac{f''(\alpha^k)}{2}(x^{k+1} - x^k)^2.$$

Nyní použijeme věty o střední hodnotě ve tvaru ($f(\xi) = 0$)

$$f(x^{k+1}) = f(x^{k+1}) - f(\xi) = f'(\beta^k)(x^{k+1} - \xi),$$

β^k leží mezi body x^{k+1} a ξ. Dosazením do předchozího vztahu odtud ihned plyne

$$\left|x^{k+1} - \xi\right| = \frac{\left|f''(\alpha^k)\right|}{2\left|f'(\beta^k)\right|}(x^{k+1} - x^k)^2,$$

a tedy i (2.19).

□

Následující věta ukazuje, že za jistých podmínek lze zajistit monotonní konvergenci posloupnosti generované Newtonovou metodou. Tyto podmínky se někdy nazývají *Fourierovy podmínky*.

Věta 2.10. *Nechť $f \in C^2[a,b]$ a nechť rovnice $f(x) = 0$ má v intervalu jediný kořen ξ. Nechť f', f'' nemění znaménka na intervalu $[a,b]$, přičemž $f'(x) \neq 0$, $\forall x \in [a,b]$. Nechť počáteční aproximace x^0 je ten z krajních bodů a, b, v němž znaménko funkce je stejné jako znaménko f'' na intervalu $[a,b]$. Pak posloupnost $\{x^k\}_{k=0}^{\infty}$ určená Newtonovou metodou konverguje monotonně k bodu ξ.*

Důkaz. Vyšetříme případ $f(a) < 0$, $f(b) > 0$, $f'(x) > 0$, $f''(x) \geq 0$, $\forall x \in [a,b]$. V ostatních případech je důkaz obdobný.

Newtonova metoda je tvaru

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}, \qquad k = 0, 1, 2, \ldots$$

Důkaz provedeme indukcí. Zvolme podle předpokladu za počáteční aproximaci bod $b = x^0$. Je třeba ukázat, že $\xi \leq x^1 < x^0$.

Jelikož

$$x^1 = x^0 - \frac{f(x^0)}{f'(x^0)}$$

a $f(x^0) > 0$, $f'(x^0) > 0$, plyne odtud, že $x^1 < x^0$.

Dále užijeme Taylorova vzorce:

$$0 = f(\xi) = f(x^0) + f'(x^0)(\xi - x^0) + \frac{f''(\eta^0)}{2}(\xi - x^0)^2, \quad \eta^0 \in (\xi, x^0).$$

Protože $f''(\eta^0)(\xi - x^0)^2/2 \geq 0$, plyne z tohoto vztahu, že

$$f(x^0) + f'(x^0)(\xi - x^0) \leq 0,$$

neboť součet členů v předchozím vztahu se rovná nule. Odtud ale plyne

$$\xi \leq x^0 - \frac{f(x^0)}{f'(x^0)} = x^1.$$

To znamená, že

$$\xi \leq x^1 < x^0.$$

Nyní za předpokladu, že platí

$$\xi \leq x^k < x^{k-1} < \ldots < x^0$$

se stejným způsobem ukáže, že

$$\xi \leq x^{k+1} < x^k < \ldots < x^0.$$

Posloupnost $\{x^k\}_{k=0}^{\infty}$ konverguje monotonně k bodu ξ. □

V předchozím příkladu jsme zvolili počáteční aproximaci v souladu s těmito podmínkami.

Poznámka 6. Nevhodná volba počáteční aproximace pro Newtonovu metodu může vést ke zcela chybným výsledkům, jak ukazuje následující příklad.

Příklad 2.9. Nechť je dáno číslo $a > 0$. Je třeba vypočítat převrácenou hodnotu tohoto čísla bez použití dělení. Najděte vhodnou funkci f a použijte Newtonovu metodu.
Řešení. Vhodná funkce f je tvaru $f(x) = 1/x - a$.

Newtonova funkce je totiž v tomto případě tvaru

$$x^{k+1} = x^k - \frac{\frac{1}{x^k} - a}{-\frac{1}{(x^k)^2}} = 2x^k - a(x^k)^2 = x^k(2 - ax^k).$$

Zabývejme se nyní konkrétní úlohou a zvolme $a = 10$. Separujme kořen rovnice $f(x) = 1/x - 10$.

Je zřejmě

$$f(0{,}01) > 0,$$

$$f(1) < 0.$$

Kořen leží v intervalu $[0{,}01; 1]$. Protože $f'(x) = -1/x^2 < 0$ pro $x \in [0{,}01; 1]$, je tento kořen jediný.

Dále $f''(x) = 2/x^3 > 0$, $x \in [0{,}01; 1]$. Vhodná počáteční aproximace je tedy $x^0 = 0{,}01$. Výpočet probíhá takto:

$$\begin{array}{llll} x^0 = 0{,}01 & x^2 = 0{,}03439 & \cdots & x^7 = 0{,}099882 \\ x^1 = 0{,}019 & x^3 = 0{,}08147 & & \end{array}$$

Pro tuto počáteční aproximaci posloupnost $\{x^k\}$ konverguje monotonně k bodu $\xi = 0{,}1$.

Zvolme nyní počáteční aproximaci $x^0 = 1$. Jednotlivé iterace jsou:

$$\begin{array}{ll} x^0 = 1 & x^2 = -656 \\ x^1 = -8 & x^3 = -4304672 \end{array}$$

Počáteční aproximace $x^0 = 1$ je špatnou počáteční aproximací. Doporučujeme čtenáři, aby sestrojil graf funkce $f(x) = 1/x - 10$, příslušné tečny a jejich průsečíky s osou x!

Následující příklad rovněž ilustruje „zajímavé“ chování iterační posloupnosti pro různé počáteční aproximace.

Příklad 2.10. Řešme rovnici $\operatorname{arctg} x = 0$. Je zřejmé, že kořen $\xi \in [a, b]$, $a < 0$, $b > 0$, neboť $\operatorname{arctg} a < 0$, $\operatorname{arctg} b > 0$. Dále

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

Víme, že kořen $\xi = 0$, ale vyšetřme tento případ podrobněji. První derivace je stále kladná, druhá derivace mění znaménko v bodě 0. Iterační funkce pro Newtonovu metodu je tvaru

$$g(x) = x - (1+x^2)\operatorname{arctg} x$$

a Newtonova iterační metoda

$$x^{k+1} = x^k - \left(1 + (x^k)^2\right)\operatorname{arctg} x^k.$$

Zvolme počáteční aproximaci $x^0 = 1{,}5$. Výsledné iterace jsou

k	x^k	$f(x^k)$
1	1,5	0,982793723
2	−1,6940796	−1,037546359
3	2,321126961	1,164002042
4	−5,114087837	−1,377694529

Zřejmě tato posloupnost nekonverguje ke kořenu $\xi = 0$. Tuto skutečnost lze objasnit takto (viz obr. 2.15):

Je $g'(x) = -2x \operatorname{arctg} x$. Na intervalu $[-1{,}5; 1{,}5]$ není splněna podmínka $|g'(x)| \leq q < 1$, a tedy není zaručena konvergence posloupnosti $\{x^k\}$.

Uvažujme nyní interval $[-0{,}75; 0{,}75]$. Funkce g zobrazuje tento interval do sebe a $|g'(x)| \leq q < 1$ na tomto intervalu. Podle věty 2.5 posloupnost určená iterační metodou bude konvergovat pro každou počáteční aproximaci $x^0 \in [-0{,}75; 0{,}75]$. Zvolme tedy $x^0 = 0{,}75$. Posloupnost iterací

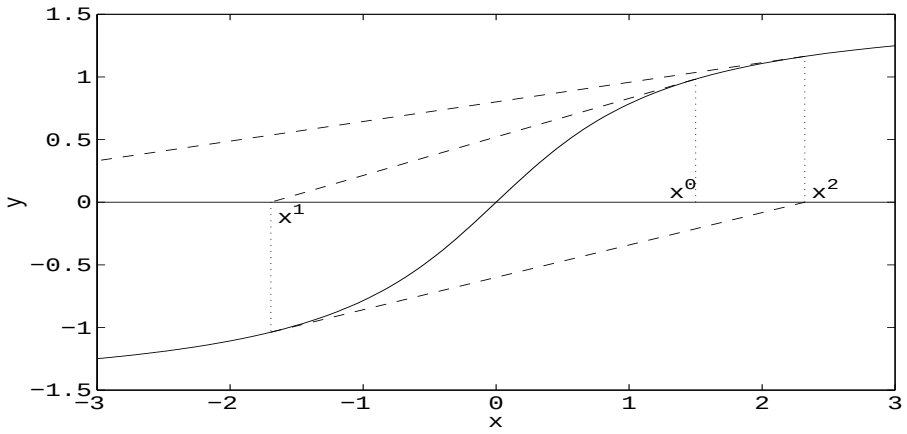
k	x^k	$f(x^k)$
1	0,75	0,643501109
2	−0,255470482	−0,250120688
3	0,010974374	0,010973934
4	$-8{,}81125.10^{-7}$	$-8{,}81125.10^{-7}$

konverguje k bodu $\xi = 0$ (viz obr. 2.16).

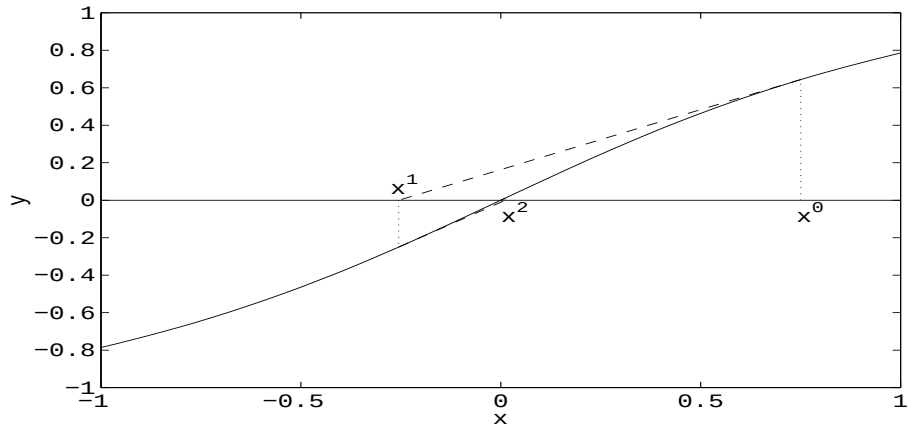
§ 2.5. Metoda sečen

Výpočet iterací $\{x^k\}_{k=0}^{\infty}$ pomocí Newtonovy metody požaduje na každém kroku výpočet $f'(x^k)$. Někdy může být tento výpočet náročný a z tohoto důvodu aproximujeme první derivaci diferencí

$$f'(x^k) \approx \frac{f(x^k) - f(x^{k-1})}{x^k - x^{k-1}}, \qquad k = 1, 2, \ldots$$



Obr. 2.15: Newtonova metoda pro $f(x) = \operatorname{arctg}(x)$, $x^0 = 1{,}5$



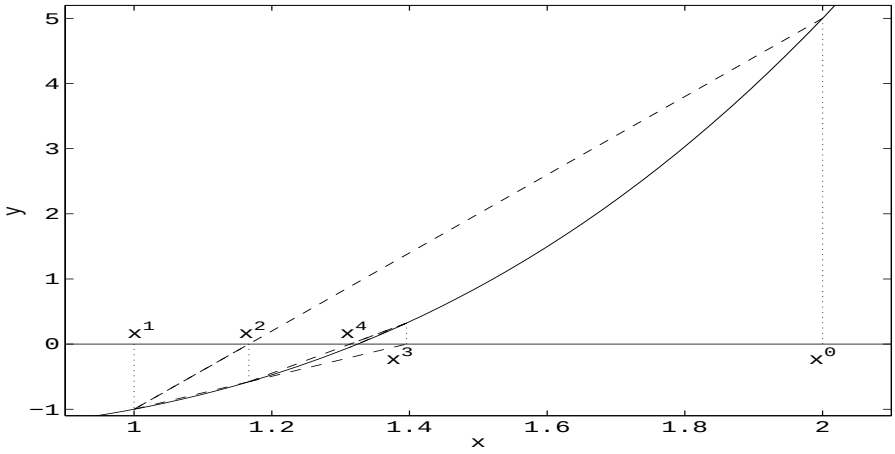
Obr. 2.16: Newtonova metoda pro $f(x) = \operatorname{arctg}(x)$, $x^0 = 0{,}75$

Výsledná iterační metoda

$$x^{k+1} = x^k - \frac{x^k - x^{k-1}}{f(x^k) - f(x^{k-1})} f(x^k), \qquad k = 0, 1, \ldots \tag{2.20}$$

se nazývá *metoda sečen*. Patří mezi tzv. quasi-Newtonovy metody. Je to metoda dvoukroková, neboť pro výpočet x^{k+1} potřebujeme dvě předchozí aproximace x^k, x^{k-1}, a tedy i dvě počáteční aproximace. Geometrický význam metody (2.20) je zřejmý z obrázku 2.17. Aproximace x^{k+1} je průsečík sečny vedené body $[x^{k-1}, f(x^{k-1})]$, $[x^k, f(x^k)]$ s osou x. Tabulka udává aproximace získané metodou sečen pro funkci $f(x) = x^3 - x - 1$.

k	x^k	$f(x^k)$
0	2	5
1	1	-1
2	$1{,}1\overline{6}$	$-0{,}57870370$
3	1,39560440	0,32263052
4	$1{,}31365\overline{6}$	$-0{,}04668748$
5	1,32401612	$-0{,}00299114$
6	1,32472525	0,00003110
7	1,32471795	-2.10^{-8}



Obr. 2.17: Metoda sečen, $f(x) = x^3 - x - 1$, $x^0 = 2$, $x^1 = 1$

Je $|(x^7 - x^6)/x^7| \approx 5.10^{-6}$ a $|f(x^7)| \approx 2.10^{-8} \Rightarrow x^7$ je dobrou aproximací hledaného kořene.

Metoda sečen je metoda dvoukroková a k důkazu konvergence nelze tedy použít věty 2.5. Pojednejme nyní o konvergenci a volbě počátečních aproximací:

Věta 2.11. *Nechť rovnice $f(x) = 0$ má kořen ξ a nechť derivace f', f'' jsou spojité v okolí bodu ξ, přičemž $f'(\xi) \neq 0$. Posloupnost určená metodou sečen konverguje ke kořenu ξ, pokud zvolíme počáteční aproximace x^0, x^1 dostatečně blízko bodu ξ a metoda je řádu $(1+\sqrt{5})/2 \approx 1{,}618$.*

Důkaz. Odečteme od pravé a levé strany rovnice

$$x^{k+1} = x^k - \frac{x^k - x^{k-1}}{f(x^k) - f(x^{k-1})} f(x^k)$$

hodnotu ξ:

$$x^{k+1} - \xi = x^k - \xi - \frac{x^k - x^{k-1}}{f(x^k) - f(x^{k-1})} f(x^k).$$

Dále upravíme tento vztah takto:

$$x^{k+1} - \xi = x^k - \xi - \frac{x^k - x^{k-1}}{f(x^k) - f(x^{k-1})} \left(f(x^k) - f(\xi)\right) \frac{x^k - \xi}{x^k - \xi} =$$

$$= (x^k - \xi) \left(1 - \frac{\dfrac{f(x^k) - f(\xi)}{x^k - \xi}}{\dfrac{f(x^k) - f(x^{k-1})}{x^k - x^{k-1}}} \right) \tag{2.21}$$

Označme dále

$$f[x^k, \xi] = \frac{f(x^k) - f(\xi)}{x^k - \xi}, \qquad f[x^k, x^{k-1}] = \frac{f(x^k) - f(x^{k-1})}{x^k - x^{k-1}} \tag{2.22}$$

$$f[x^{k-1}, x^k, \xi] = \frac{f[x^k, x^{k-1}] - f[x^k, \xi]}{x^{k-1} - \xi} \tag{2.23}$$

Vztahy (2.22) a (2.23) se nazývají poměrné diference 1. resp. 2. řádu a pojednáme o nich později v kapitole 6. Lze snadno ukázat pomocí věty o střední hodnotě a Taylorova vzorce (viz též kapitola 6), že

$$f[x^k, x^{k-1}] = f'(\alpha^k),$$

kde α^k leží v intervalu určeném body x^k, x^{k-1},

$$f[x^{k-1}, x^k, \xi] = \tfrac{1}{2} f''(\beta^k),$$

kde β^k leží v intervalu určeném body x^{k-1}, x^k, ξ. Vztah (2.21) můžeme nyní psát ve tvaru

$$x^{k+1} - \xi = (x^k - \xi) \left(1 - \frac{f[x^k, \xi]}{f[x^k, x^{k-1}]} \right) =$$

$$= (x^k - \xi) \frac{f[x^k, x^{k-1}] - f[x^k, \xi]}{f[x^k, x^{k-1}]} =$$

$$= (x^k - \xi)(x^{k-1} - \xi) \frac{f[x^{k-1}, x^k, \xi]}{f[x^k, x^{k-1}]} \Rightarrow$$

$$\Rightarrow \quad x^{k+1} - \xi = (x^k - \xi)(x^{k-1} - \xi) \frac{f''(\beta^k)}{2 f'(\alpha^k)}. \tag{2.24}$$

Předpokládáme, že kořen ξ je jednoduchý. Z toho plyne, že $f'(x) \neq 0$ v okolí bodu ξ. Existuje tedy číslo M a interval $J = \{x \mid |x - \xi| \leq \varepsilon\}$ tak, že

$$\left| \frac{f''(\beta^k)}{2f'(\alpha^k)} \right| \leq M, \qquad \forall \alpha^k, \beta^k \in J.$$

Položme $e_k = M|x^k - \xi|$ a e_0, $e_1 \leq \delta$, $\delta < \min\{1, \varepsilon M\}$. Pak lze snadno indukcí dokázat, že

$$e_{k+1} \leq e_k e_{k-1}, \qquad k = 1, 2, \ldots, \tag{2.25}$$

$$e_k \leq \min\{1, \varepsilon M\}. \tag{2.26}$$

Ukážeme nyní opět indukcí, že platí

$$e_k \leq K^{q^k}, \qquad k = 0, 1, 2, \ldots, \tag{2.27}$$

kde $K = \max\left\{e_0, \sqrt[q]{e_1}\right\} < 1$ a q je kladný kořen rovnice $q^2 = q+1$, $q = (1+\sqrt{5})/2$. Je zřejmé, že pro $k = 0, 1$ vztah platí. Předpokládejme nyní, že (2.27) platí pro k a dokážeme, že platí pro $k+1$. Ze vztahu (2.25) plyne

$$e_{k+1} \leq e_k e_{k-1} \leq K^{q^k} K^{q^{k-1}} = K^{q^{k-1}(q+1)} = K^{q^{k+1}},$$

neboť $q + 1 = q^2$. Vztah (2.27) tedy platí pro všechna k. Ze vztahu (2.27) rovněž plyne, že metoda sečen konverguje alespoň tak jako metoda řádu $q = (1 + \sqrt{5})/2$. □

Poznámka 7. Jeden krok metody sečen požaduje pouze výpočet jedné funkční hodnoty. Dva kroky metody sečen jsou nejvýše tak „drahé“ jako jeden krok metody Newtonovy. Jelikož

$$K^{q^{k+2}} = \left(K^{q^k}\right)^{q^2} = \left(K^{q^k}\right)^{q+1},$$

představují dva kroky metody sečen metodu řádu $q^2 = q + 1 \approx 2{,}618$. Lze tedy říci, že *metoda sečen lokálně konverguje rychleji než metoda Newtonova*.

§ 2.6. Metoda regula falsi

Jak jsme viděli, konvergence Newtonovy metody i metody sečen závisí na vhodné volbě počáteční aproximace nebo dvou počátečních aproximací. Pro získání počátečních aproximací lze použít metodu bisekce, ale další vhodnou metodou je metoda *regula falsi*. Popíšeme nyní tuto metodu.

Předpokládejme, že $f(a)f(b) < 0$, $f \in C[a, b]$. Metoda bisekce užívá středu intervalu $[a, b]$, ale lepší aproximaci získáme, jestliže najdeme bod $c \in (a, b)$, ve kterém přímka vedená body $[a, f(a)]$, $[b, f(b)]$ protíná osu x. Další postup aplikujeme na ten z intervalů $[a, c]$, $[c, b]$, v jehož koncových bodech má funkce f opačná

znaménka. (Může samozřejmě také nastat případ $f(c) = 0$, a tedy $c = \xi$ je kořen.) Obecně lze tuto metodu, která se nazývá *metoda regula falsi*, zapsat takto:

$$x^{k+1} = x^k - \frac{x^k - x^s}{f(x^k) - f(x^s)} f(x^k), \qquad k = 0, 1, \ldots, \tag{2.28}$$

kde $s = s(k)$ je největší index takový, že $f(x^k)f(x^s) < 0$. Předpokládáme, že počáteční aproximace x^0, x^1 jsou vybrány tak, že $f(x^0)f(x^1) < 0$ (tj. např. $x^0 = a$, $x^1 = b$).

Věta 2.12. *Nechť $f \in C[a,b]$, $f(a)f(b) < 0$ a nechť ξ je jediný kořen v $[a,b]$. Pak posloupnost $\{x^k\}_{k=0}^{\infty}$ určená metodou regula falsi konverguje pro libovolné počáteční aproximace x^0, $x^1 \in [a,b]$, $f(x^0)f(x^1) < 0$, ke kořenu $\xi \in (a,b)$ funkce f a je to metoda prvního řádu.*

Důkaz je obecném případě značně obsáhlý (viz [5]).

Na tomto místě se budeme zabývat případem, kdy $f \in C[a,b]$ a f'' nemění znaménko na $[a,b]$ a popíšeme situaci z geometrického hlediska. Předpokládejme, že $f \in [a,b]$, $f(a) < 0$, $f(b) > 0$, f má jediný kořen v intervalu $[a,b]$, $f'(x) > 0$, $f''(x) \le 0$, $\forall x \in [a,b]$. Položme $a = x^0$, $b = x^1$. Z předpokladů plyne, že pro bod

$$x^2 = x^1 - \frac{x^1 - x^0}{f(x^1) - f(x^0)} f(x^1)$$

platí $x^0 < x^2 < x^1$. Funkce f je na intervalu $[x^0, x^1]$ konkávní a to znamená, že sečna určená body $[x^0, f(x^0)]$, $[x^1, f(x^1)]$ leží pod grafem funkce $y = f(x)$, a tedy průsečík x^2 této sečny s osou x leží napravo od bodu ξ: $\xi < x^2 < x^1$. To však znamená, že $f(x^2) > 0$ a metodu regula falsi aplikujeme na interval $[x^0, x^2]$:

$$x^3 = x^2 - \frac{x^2 - x^0}{f(x^2) - f(x^0)} f(x^2)$$

Indukcí plyne obecný vztah

$$x^{k+1} = x^k - \frac{x^k - x^0}{f(x^k) - f(x^0)} f(x^k), \qquad f(x^k)f(x^0) < 0, \qquad k = 1, 2, \ldots$$

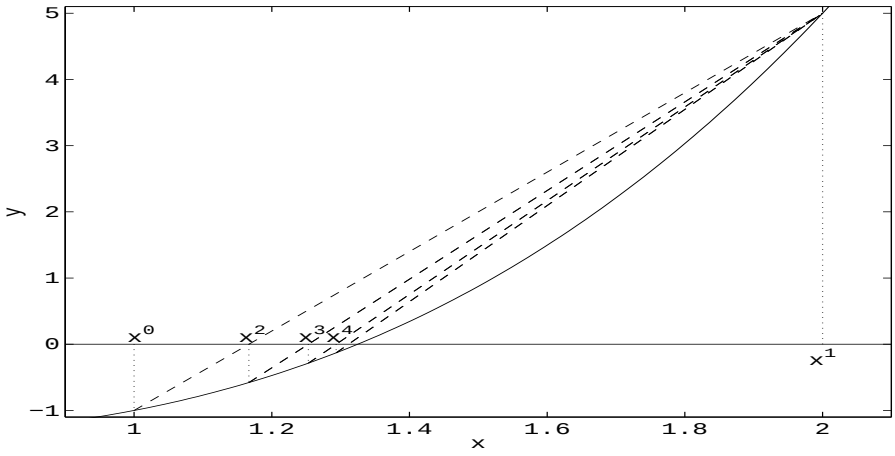
To znamená, že bod x^0 je v jistém smyslu „pevný“. Všechny sečny vycházejí z tohoto bodu. Posloupnost $\{x^k\}$ v tomto případě zřejmě monotonně konverguje k bodu $x = \xi$: $\xi \le x^k < x^{k-1} < \ldots < x^1$, $\lim_{k\to\infty} x^k = \xi$.

Obdobné úvahy lze provést i v dalších případech. Výsledek lze formulovat takto:

Nechť $f \in C[a,b]$, $f(a)f(b) < 0$, a nechť f'' nemění znaménko na intervalu $[a,b]$. Pak „pevný“ je ten koncový bod intervalu, v němž znaménko funkce je stejné jako znaménko f'' na intervalu $[a,b]$ a posloupnost $\{x^k\}$ konverguje monotonně ke kořenu $x = \xi$.

Podrobný důkaz lze provést užitím Taylorova vzorce.

k	x^k	$f(x^k)$
0	1	-1
1	2	5
2	$1{,}1\overline{6}$	$-0{,}57870370$
3	1,25311203	$-0{,}28536303$
4	1,29343740	$-0{,}12954209$
5	1,31128102	$-0{,}05658849$
6	1,31898850	$-0{,}02430375$
7	1,32228272	$-0{,}01036185$
8	1,32368429	$-0{,}00440395$



Obr. 2.18: Metoda regula falsi, $f(x) = x^3 - x - 1$, $x^0 = 1$, $x^1 = 2$

Metoda regula falsi pro funkci $f(x) = x^3 - x - 1$ je ilustrována na obr. 2.18. Jelikož $f'(x) = 3x^2 - 1$ a $f''(x) > 0$ pro $x \in [1, 2]$, je bod $b = 2$ „pevný“. Tabulka udává aproximace získané metodou regula falsi pro danou funkci.

Užijeme kriterií (2.4), (2.5) pro odhad chyby. Je $|(x^8 - x^7)/x^7| \approx 0{,}0014$ a $f(x^8) \approx 0{,}0044$. Můžeme tedy říci, že iterace x^8 aproximuje kořen ξ s chybou menší než 5.10^{-3}.

§ 2.7. Quasi Newtonova metoda

Hlavní myšlenkou této metody je nahrazení tečny použité v Newtonově metodě sečnou procházející bodem $(x^k, f(x^k))$ a bodem $(x^k + f(x^k), f(x^k + f(x^k)))$, respektive bodem $(x^k - f(x^k), f(x^k - f(x^k)))$. Přitom pokud je bod x^k blízko hledaného kořene ξ, pak hodnota $f(x^k)$ je blízká nule a sečna procházející uvedenými body je

blízká tečně vedené bodem x^k. Jedná se tedy o metodu blízkou metodě Newtonově, zde má také původ název této metody.

Nahrazením $f'(x^k)$ ve vztahu pro Newtonovu metodu (2.16) přibližnou hodnotou

$$\frac{f(x^k) - f(x^k \pm f(x^k))}{x^k - (x^k \pm f(x^k))} = \frac{f(x^k) - f(x^k \pm f(x^k))}{\mp f(x^k)}$$

dostáváme iterační vztah

$$x^{k+1} = x^k - f(x^k)\frac{\mp f(x^k)}{f(x^k) - f(x^k \pm f(x^k))} = x^k \pm \frac{f^2(x^k)}{f(x^k) - f(x^k \pm f(x^k))} \quad (2.29)$$

a iterační funkce má tedy tvar

$$g(x) = x \pm \frac{f^2(x)}{f(x) - f(x \pm f(x))}\,. \quad (2.30)$$

Věta 2.13. *Nechť $f \in C^1[a,b]$, $\xi \in [a,b]$ nechť je řešením rovnice $f(x) = 0$ a $f'(\xi) \neq 0$. Pak existuje $\varepsilon > 0$ tak, že posloupnost $\{x^k\}_{k=0}^{\infty}$ generovaná quasi Newtonovou metodou konverguje k bodu ξ pro každou počáteční aproximaci $x^0 \in [\xi - \varepsilon, \xi + \varepsilon] \cap [a,b]$. Pokud má funkce f v okolí bodu ξ spojitou druhou derivaci, je řád metody alespoň 2.*

Důkaz. Důkaz provedeme pro $\xi \in (a,b)$. V případě, že ξ je jedním z krajních bodů intervalu, úvahy jsou obdobné při použití jednostranných intervalů.
Nechť δ je takové, že $[\xi - \delta, \xi + \delta] \subseteq [a,b]$. Pak existuje δ_0, že pro $x \in [\xi - \delta_0, \xi + \delta_0]$ platí $|f(x)| \leq \delta/2$. Položme $\delta_1 = \min\{\delta_0, \delta/2\}$. Potom pro $x \in [\xi - \delta_1, \xi + \delta_1]$ dostáváme

$$|\xi - (x \pm f(x))| \leq |\xi - x| + |f(x)| \leq \delta/2 + \delta/2 = \delta$$

takže bod $x \pm f(x)$ leží v intervalu $[\xi - \delta, \xi + \delta] \subseteq [a,b]$ a je tedy definována hodnota $f(x \pm f(x))$.

Dále pomocí l'Hospitalova pravidla spočítáme limitu

$$\lim_{x\to\xi} \frac{\mp f(x)}{f(x) - f(x \pm f(x))} = \lim_{x\to\xi} \frac{\mp f'(x)}{f'(x) - f'(x \pm f(x))(1 \pm f'(x))} =$$

$$= \lim_{x\to\xi} \frac{\mp f'(x)}{f'(x) - f'(x)(1 \pm f'(x))} = \lim_{x\to\xi} \frac{\mp f'(x)}{f'(x) - f'(x) \mp (f'(x))^2} = \frac{1}{f'(\xi)}\,,$$

protože $f(x) \to 0$ pro $x \to \xi$.
Odtud dostáváme

$$g(\xi) = \lim_{x\to\xi} g(x) = \xi - f(\xi) \lim_{x\to\xi} \frac{\mp f(x)}{f(x) - f(x \pm f(x))} = \xi - \frac{f(\xi)}{f'(\xi)} = \xi\,,$$

takže ξ je pevným bodem funkce g. Navíc

$$g'(\xi) \;=\; \lim_{x\to\xi} \frac{g(x) - g(\xi)}{x - \xi} = \lim_{x\to\xi} \frac{1}{x - \xi}\left(x - f(x)\frac{\mp f(x)}{f(x) - f(x \pm f(x))} - \xi\right) =$$

$$= \lim_{x\to\xi}\frac{x-\xi}{x-\xi} - \lim_{x\to\xi}\frac{f(x)}{x-\xi}\cdot\frac{\mp f(x)}{f(x)-f(x\pm f(x))} =$$
$$= 1 - \lim_{x\to\xi}\frac{f(x)-f(\xi)}{x-\xi}\cdot\lim_{x\to\xi}\frac{\mp f(x)}{f(x)-f(x\pm f(x))} = 1 - f'(x)\frac{1}{f'(x)} = 0\,.$$

Protože f má spojitou derivaci v ξ, má v tomto bodě spojitou derivaci i funkce g, a proto existuje $\varepsilon > 0$ takové, že $[\xi-\varepsilon,\xi+\varepsilon] \subseteq [\xi-\delta_1,\xi+\delta_1]$ a pro $x \in [\xi-\varepsilon,\xi+\varepsilon]$ je $|g(x)| \le q < 1$. Interval $[\xi-\varepsilon,\xi+\varepsilon]$ funkce g zobrazuje do sebe, neboť jestliže $x \in [\xi-\varepsilon,\xi+\varepsilon]$, pak

$$|g(x)-\xi| = |g(x)-g(\xi)| \le q|x-\xi| < |x-\xi| \le \varepsilon\,.$$

Podle věty 2.5 konverguje metoda pro libovolnou počáteční aproximaci $x^0 \in [\xi-\varepsilon,\xi+\varepsilon]$. Pokud má f v okolí bodu ξ spojitou druhou derivaci, má zde spojitou druhou derivaci i funkce g, a jelikož platí $g'(\xi) = 0$, podle věty 2.4 je quasi Newtonova metoda řádu alespoň 2. □

Příklad 2.11. Použijeme quasi Newtonovou metodou na stejnou úlohu jako v příkladě 2.8, tj. pro nalezení kořene funkce $f(x) = x^3 - x - 1$ ležící v intervalu $[1,2]$.
Řešení. Quasi Newtonova metoda (varianta +) dává iterační vztah

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

Je samozřejmě možné výraz upravit dosazením dané funkce, výsledný vztah je ovšem poměrně komplikovaný. Při vlastním výpočtu je lepší použít např. substituci $y^k = f(x^k)$ a pak použít vztah

$$x^{k+1} = x^k + \frac{(y^k)^2}{y^k - f(x^k+y^k)}.$$

Za počáteční aproximaci zvolme $x^0 = 2$. V tabulce jsou uvedeny jednotlivé aproximace a funkční hodnoty

k	x^k	$f(x^k)$
0	1,4	0,344000000
1	1,346609850	0,095276011
2	1,326900496	0,009326670
3	1,324741149	0,000098908
4	1,324717960	0,000000011
5	1,324717957	$2{,}2204\cdot10^{-16}$

Vidíme, že je potřeba poměrně velmi přesnou počáteční iteraci, aby hodnota $x^0 + f(x^0)$ ležela v daném intervalu, ale nemusíme počítat derivace.

Poznámka 8. Dosud probrané metody patří mezi nejužívanější metody k řešení nelineárních rovnic. Z nich metoda bisekce a regula falsi patří mezi *vždy konvergentní metody*, neboť posloupnost aproximací generovaná těmito metodami vždy

konverguje k hledanému kořenu spojité funkce na daném intervalu. Jejich nevýhodou je pomalá konvergence, lze jich ale s výhodou použít pro nalezení dobré počáteční aproximace pro některou jinou metodu, která konverguje rychleji.

§ 2.8. Iterační metody pro násobné kořeny

Doposud jsme předpokládali, že kořen ξ je jednoduchým kořenem rovnice $f(x) = 0$. Dá se ukázat, viz např. [5], [13], že uvedené iterační metody konvergují lineárně, má-li hledaný kořen násobnost M větší než 1. Ale známe-li násobnost kořene, můžeme modifikovat Newtonovu metodu tak, že konvergence bude opět kvadratická.

Věta 2.14. *Nechť kořen ξ má násobnost $M > 1$. Pak modifikovaná Newtonova metoda*

$$x^{k+1} = x^k - M\frac{f(x^k)}{f'(x^k)} \tag{2.31}$$

je metoda druhého řádu.

Důkaz. Nechť

$$x^{k+1} = x^k - M\frac{f(x^k)}{f'(x^k)}.$$

Pak

$$\xi - x^{k+1} = \xi - x^k + M\frac{f(x^k)}{f'(x^k)} \tag{2.32}$$

a odtud

$$(\xi - x^{k+1})f'(x^k) = \sigma(x^k), \tag{2.33}$$

$$\sigma(x) = (\xi - x)f'(x) + Mf(x).$$

Derivováním dostaneme

$$\sigma^{(j)}(x) = Mf^{(j)}(x) + (\xi - x)f^{(j+1)}(x) - jf^{(j)}(x).$$

Bod ξ je M-násobným kořenem funkce f ($f^{(j)}(\xi) = 0$, $j = 0, 1, \ldots, M-1$, $f^{(M)}(\xi) \neq 0$) a tedy

$$\sigma^{(j)}(\xi) = 0, \qquad j = 0, 1, \ldots, M, \qquad \sigma^{(M+1)}(\xi) \neq 0.$$

Aplikací Taylorova vzorce pro funkci σ odtud plyne

$$\sigma(x) = \frac{(x-\xi)^{M+1}}{(M+1)!}\sigma^{(M+1)}(\alpha_1). \tag{2.34}$$

Na druhé straně (opět z Taylorova vzorce)

$$f'(x) = \frac{(x-\xi)^{M-1}}{(M-1)!}f^{(M)}(\alpha_2). \tag{2.35}$$

Dosazením vztahů (2.34) a (2.35) do (2.33) dostaneme

$$(\xi - x^{k+1})\frac{(x^k - \xi)^{M-1}}{(M-1)!} f^{(M)}(\alpha_2^k) = \frac{(x^k - \xi)^{M+1}}{(M+1)!}\sigma^{(M+1)}(\alpha_1^k)$$

a odtud

$$\left(\xi - x^{k+1}\right) = \frac{\left(\xi - x^k\right)^2}{M(M+1)} \frac{\left(\sigma^{(M+1)}(\alpha_1^k)\right)}{\left(f^{(M)}(\alpha_2^k)\right)}.$$

Metoda je tedy řádu 2, neboť $\left|f^{(M)}(x)\right| \geq m > 0$ v okolí bodu $x = \xi$ a $\sigma^{(M+1)}(\xi) \neq 0$. □

Poznámka 9. „Klasická“ Newtonova metoda konverguje pro násobný kořen lineárně.

Následující tabulka ukazuje srovnání rychlostí konvergence pro jednotlivé metody. Zde $e_k = \left|x^k - \xi\right|$.

Metoda	**Speciální případy**	**Posloupnost chybových členů**
Bisekce		$e_{k+1} \approx \frac{1}{2}\, e_k$
Regula falsi		$e_{k+1} \approx C_R\, e_k$
Metoda sečen	násobný kořen	$e_{k+1} \approx C_{SN}\, e_k$
Newtonova metoda	násobný kořen	$e_{k+1} \approx C_{NN}\, e_k$
Metoda sečen	jednoduchý kořen	$e_{k+1} \approx C_S\, e_k^{1.618}$
Newtonova metoda	jednoduchý kořen	$e_{k+1} \approx C_N\, e_k^2$
quasi Newtonova metoda	jednoduchý kořen	$e_{k+1} \approx C_Q\, e_k^2$
Modifikovaná Newtonova metoda	násobný kořen	$e_{k+1} \approx C_{MN}\, e_k^2$

Poznámka 10. Obvykle násobnost kořene předem neznáme. Víme ale, že funkce $u(x) = f(x)/f'(x)$ má v bodě $x = \xi$ jednoduchý kořen bez ohledu na násobnost kořene původní funkce. Místo rovnice $f(x) = 0$ uvažujme tedy rovnici

$$u(x) = 0,$$

jejíž kořeny jsou totožné s kořeny dané rovnice a jsou všechny jednoduché. Nyní můžeme aplikovat výše uvedené metody na tuto funkci a řád konvergence metody se nezmění.

§ 2.9. Urychlení konvergence

Nyní vyložíme techniku nazývanou *Aitkenovou δ^2-metodou*, která může být použita k urychlení konvergence libovolné lineárně konvergentní posloupnosti nezávisle na tom, jak je tato posloupnost generována.

Věta 2.15. (Aitkenova δ^2-metoda). *Nechť je dána posloupnost $\{x^k\}_{k=0}^{\infty}$, $x^k \neq \xi$, $k = 0, 1, 2, \ldots$, $\lim_{k\to\infty} x^k = \xi$, a nechť tato posloupnost splňuje podmínky*

$$x^{k+1} - \xi = (C + \gamma_k)(x^k - \xi), \quad k = 0, 1, 2, \ldots \qquad |C| < 1, \qquad \lim_{k\to\infty} \gamma_k = 0.$$

Pak posloupnost

$$\hat{x}^k = x^k - \frac{(x^{k+1} - x^k)^2}{x^{k+2} - 2x^{k+1} + x^k} \tag{2.36}$$

je definována pro všechna dostatečně velká k a platí

$$\lim_{k\to\infty} \frac{\hat{x}^k - \xi}{x^k - \xi} = 0,$$

tj. posloupnost $\{\hat{x}^k\}$ konverguje k limitě ξ rychleji než posloupnost $\{x^k\}$.

Důkaz. Ověříme, zda posloupnost $\{\hat{x}^k\}$ je definována pro dostatečně velká k. Počítejme

$$\begin{aligned} x^{k+2} - 2x^{k+1} + x^k &= (x^{k+2} - \xi) - 2(x^{k+1} - \xi) + (x^k - \xi) = \\ &= (x^{k+1} - \xi)(C + \gamma_{k+1}) - 2(x^k - \xi)(C + \gamma_k) + (x^k - \xi) = \\ &= (x^k - \xi)(C + \gamma_k)(C + \gamma_{k+1}) - 2(x^k - \xi)(C + \gamma_k) + (x^k - \xi) = \\ &= (x^k - \xi)(C^2 - 2C + 1 + \tau_k) = (x^k - \xi)\left((C-1)^2 + \tau_k\right), \end{aligned}$$

kde $\lim_{k\to\infty} \tau_k = 0$.

Pro dostatečně velká k je $x^{k+2} - 2x^{k+1} + x^k \neq 0$ a posloupnost $\{\hat{x}^k\}$ je definována.

Nyní

$$\hat{x}^k - \xi = x^k - \xi - \frac{(x^{k+1} - x^k)^2}{x^{k+2} - 2x^{k+1} + x^k} = (x^k - \xi) - \frac{\left(x^{k+1} - \xi - (x^k - \xi)\right)^2}{(x^k - \xi)\left((C-1)^2 + \tau_k\right)} =$$

$$= (x^k - \xi) - \frac{(x^k - \xi)^2(C - 1 + \gamma_k)^2}{(x^k - \xi)\left((C-1)^2 + \tau_k\right)} = (x^k - \xi)\left(1 - \frac{(C - 1 + \gamma_k)^2}{(C-1)^2 + \tau_k}\right)$$

a odtud

$$\lim_{k\to\infty} \frac{\hat{x}^k - \xi}{x^k - \xi} = \lim_{k\to\infty} \left(1 - \frac{(C - 1 + \gamma_k)^2}{(C-1)^2 + \tau_k}\right) = 0.$$

□

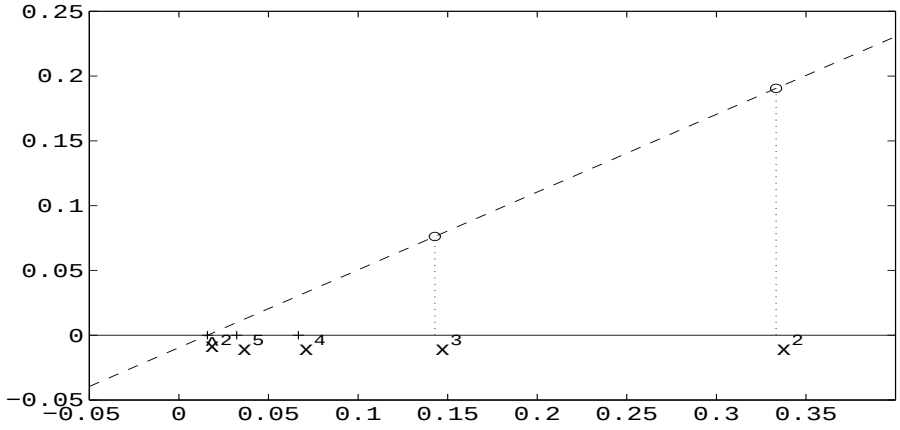
Poznámka 11. Výraz ve jmenovateli $x^{k+2} - 2x^{k+1} + x^k$ se nazývá někdy druhá centrální diference a označuje se $\delta^2 x^k$. Odtud název *Aitkenova δ^2-metoda*.

Uvedená metoda má zajímavý geometrický význam. Definujme funkci chyby ε takto:

$$\varepsilon(x^k) = x^k - x^{k+1}, \qquad \varepsilon(x^{k+1}) = x^{k+1} - x^{k+2}$$

Chceme sestrojit takovou posloupnost, která by konvergovala rychleji k bodu ξ.

Body o souřadnicích $\left[x^k, \varepsilon(x^k)\right]$, $\left[x^{k+1}, \varepsilon(x^{k+1})\right]$ vedeme přímku a její průsečík s osou x vezmeme za další aproximaci bodu ξ, tj. provedeme „extrapolaci“ (viz obr. 2.19).



Obr. 2.19: Aitkenova metoda, $x^n = 1/(2^n - 1)$

Rovnice přímky je tvaru

$$y - \varepsilon(x^k) = \frac{\varepsilon(x^k) - \varepsilon(x^{k+1})}{x^k - x^{k+1}}(x - x^k)$$

Odtud je zřejmé, že průsečík s osou x ($y = 0$) je právě hodnota $\hat{x}^k$

$$\hat{x}^k = x^k - \frac{\varepsilon(x^k)(x^k - x^{k+1})}{\varepsilon(x^k) - \varepsilon(x^{k+1})} = x^k - \frac{(x^{k+1} - x^k)^2}{x^{k+2} - 2x^{k+1} + x^k}.$$

§ 2.10. Steffensenova metoda

Vraťme se nyní k iterační metodě $x^{k+1} = g(x^k)$. Můžeme užít Aitkenovy δ^2-metody ke konstrukci posloupnosti $\{\hat{x}^k\}$, která konverguje rychleji než původní posloupnost $\{x^k\}$. Je vhodné sestrojit novou posloupnost takto:

Položme

$$y^k = g(x^k), \qquad z^k = g(y^k),$$

$$x^{k+1} = x^k - \frac{(y^k - x^k)^2}{z^k - 2y^k + x^k}. \tag{2.37}$$

V tomto případě je tedy $\varepsilon(x^k) = x^k - y^k$, $\varepsilon(y^k) = y^k - z^k$.

Posloupnost (2.37) je posloupnost sestrojená Aitkenovou δ^2-metodou. Tato iterační metoda se nazývá *Steffensenova* a může být popsána iterační funkcí φ:

$$x^{k+1} = \varphi(x^k), \tag{2.38}$$

kde

$$\varphi(x) = \frac{xg(g(x)) - g^2(x)}{g(g(x)) - 2g(x) + x}, \qquad g^2(x) = (g(x))^2. \tag{2.39}$$

Snadno lze ověřit, že iterační proces (2.38) generuje posloupnost danou vztahem (2.37). Počítejme hodnotu $\varphi(x^k)$:

$$\varphi(x^k) = \frac{x^k g(g(x^k)) - g^2(x^k)}{g(g(x^k)) - 2g(x^k) + x^k} = \frac{x^k g(y^k) - (y^k)^2}{g(y^k) - 2y^k + x^k} =$$

$$= \frac{x^k z^k - (y^k)^2}{z^k - 2y^k + x^k} = \frac{x^k(z^k - 2y^k + x^k) - (y^k - x^k)^2}{z^k - 2y^k + x^k} = x^k - \frac{(y^k - x^k)^2}{z^k - 2y^k + x^k}$$

Pro pevné body funkcí φ, g platí následující věta.

Věta 2.16.

1. *$\varphi(\xi) = \xi$ implikuje $g(\xi) = \xi$.*
2. *Jestliže $g(\xi) = \xi$, $g'(\xi)$ existuje a $g'(\xi) \neq 1$, pak $\varphi(\xi) = \xi$.*

Důkaz.

1. Z definice funkce φ dané vztahem (2.39) plyne

   $$(\xi - \varphi(\xi))(g(g(\xi)) - 2g(\xi) + \xi) = (\xi - g(\xi))^2.$$

   Tedy $\varphi(\xi) = \xi$ implikuje $g(\xi) = \xi$.

2. Předpokládejme nyní, že $\xi = g(\xi)$, g je diferencovatelná pro $x = \xi$ a $g'(\xi) \neq 1$. Pro výpočet hodnoty $\varphi(\xi)$ použijeme l'Hospitalova pravidla:

   $$\varphi(\xi) = \frac{g(g(\xi)) + \xi g'(g(\xi))g'(\xi) - 2g(\xi)g'(\xi)}{g'(g(\xi))g'(\xi) - 2g'(\xi) + 1} = \frac{\xi + \xi g'^2(\xi) - 2\xi g'(\xi)}{1 + g'^2(\xi) - 2g'(\xi)} = \xi$$

□

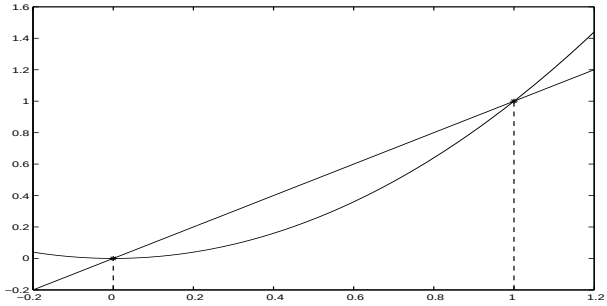
Věta 2.17. *Nechť funkce g má spojité derivace až do řádu $p + 1$ včetně v okolí bodu $x = \xi$. Nechť iterační metoda $x^{k+1} = g(x^k)$ je řádu p pro bod ξ.*

Pak pro $p > 1$ je iterační metoda $x^{k+1} = \varphi(x^k)$ řádu $2p - 1$. Pro $p = 1$ je tato metoda řádu alespoň 2 za předpokladu $g'(\xi) \neq 1$.

Důkaz lze nalézt v [5].

Poznamenejme, že funkce φ je iterační funkcí pro metodu druhého řádu, tj. metoda je lokálně kvadraticky konvergentní i v případě, že $|g'(\xi)| > 1$ a metoda $x^{k+1} = g(x^k)$ diverguje jako posloupnost.

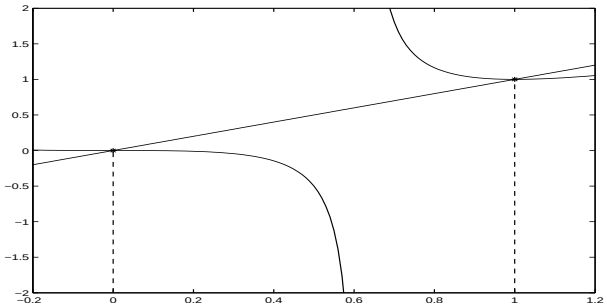
Příklad 2.12. Iterační funkce $g(x) = x^2$ má dv pevné body $\xi_1 = 0$, $\xi_2 = 1$.



Je $g'(x) = 2x$, $g''(x) = 2$, $\Rightarrow g'(\xi_1) = 0$, $g'(\xi_2) = 2$. ξ_1 je přitahujícím pevným bodem, a protože $g'(\xi_1) = 0$, je pro tento bod metoda řádu druhého. o znamená, že pro každou počáteční iteraci $x^0 \in [-1/4, 1/4]$ bude posloupnost $x^{k+1} = (x^k)^2$ konvergovat kvadraticky k bodu $\xi_1 = 0$. Bod ξ_2 je odpuzujícím pevným bodem a posloupnost $x^{k+1} = (x^k)^2$ nekonverguje k bodu ξ_2 (pro $|x^0| < 1$ posloupnost konverguje k bodu ξ_1, pro $|x^0| > 1$ diverguje). Protože $g'(\xi_i) \neq 1$, $i = 1, 2$, můžeme sestrojit iterační funkci pro Steffensenovu metodu a tato iterační funkce bude mít stejné pevné body jako funkce g.
Je

$$\begin{aligned}\varphi(x) &= \frac{x\,g(g(x)) - g^2(x)}{g(g(x)) - 2g(x) + x} = \frac{x^5 - x^4}{x^4 - 2x^2 + x} = \\ &= \frac{x^4(x-1)}{x(x-1)(x^2+x-1)} = \frac{x^3}{x^2+x-1} = \frac{x^3}{(x-\lambda_1)(x-\lambda_2)},\end{aligned}$$

kde $\lambda_{1,2} = (-1 \pm \sqrt{5})/2$.
Tato iterační funkce je vhodná pro oba pevné body ξ_1, ξ_2 a platí
$\varphi'(\xi_1) = 0$, $\varphi''(\xi_1) = 0 \Rightarrow$ metoda je řádu 3 pro bod $\xi_1 = 0$,
$\varphi'(\xi_2) = 0$, $\varphi''(\xi_2) \neq 0 \Rightarrow$ metoda je řádu 2 pro bod $\xi_2 = 1$.

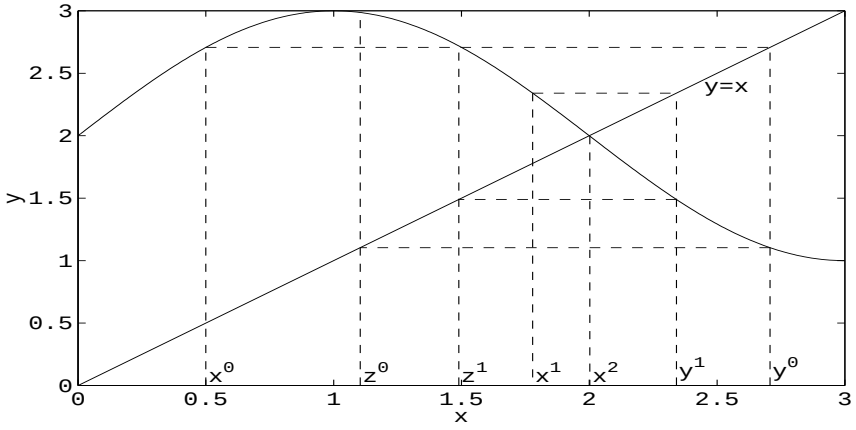


Graf funkce φ.

Jednotlivé iterace pro ξ_1: $x^0 = \frac{1}{4}$, $x^1 = -\frac{1}{44}$, $x^2 = \frac{1}{87076} \doteq 0,00001484$.
Jednotlivé iterace pro ξ_2: $x^0 = 2$, $x^1 = 1,6$, $x^2 \doteq 1,2962$, $x^3 \doteq 1,1019$.

Příklad 2.13. Aplikujme Steffensenovu metodu na iterační funkci $g(x) = \sin \frac{\pi}{2}x + 2$, jejíž pevný bod $\xi = 2$. Metoda je znázorněna na obr. 2.20.

k	x^k	y^k	z^k
0	0,50000000000000	2,70710678118655	1,10398106407319
1	1,77848375282432	2,34097786767556	1,48963705888740
2	2,00227199284588	1,99643116955900	2,00560587638570
3	1,99999999735784	2,00000000415030	1,99999999348073
4	2,00000000000000	2,00000000000000	2,00000000000000



Obr. 2.20: Steffensenova metoda (pro iterační metodu), $x = \sin(\frac{\pi}{2}x) + 2$

Poznámka 12. Hledáme-li řešení rovnice $f(x) = 0$ a iterační funkci g definujeme vztahem $g(x) = x + f(x)$ nebo $g(x) = x - f(x)$, dostaneme pomocí Steffensenovy metody

$$y^k = g(x^k) = x^k \pm f(x^k), \quad z^k = g(y^k) = y^k \pm f(y^k) = x^k \pm f(x^k) \pm f(x^k \pm f(x^k)),$$

$$\begin{aligned} x^{k+1} &= x^k - \frac{(y^k - x^k)^2}{z^k - 2y^k + x^k} = \\ &= x^k - \frac{(x^k \pm f(x^k) - x^k)^2}{x^k \pm f(x^k) \pm f(x^k \pm f(x^k)) - 2x^k \mp 2f(x^k) + x^k} = \\ &= x^k - \frac{f^2(x^k)}{\pm f(x^k \pm f(x^k)) \mp f(x^k)} = \end{aligned}$$

$$= x^k - \frac{f^2(x^k)}{\mp[f(x^k) - f(x^k \pm f(x^k))]} =$$
$$= x^k \pm \frac{f^2(x^k)}{f(x^k) - f(x^k \pm f(x^k))}.$$

V tomto případě se tedy jedná o quasi Newtonovu metodu (viz odstavec 2.7).

§ 2.11. Müllerova metoda

Na závěr této kapitoly se zmíníme ještě o Müllerově metodě. Tato metoda byla navržena v roce 1956 D. M. Müllerem. Je vhodná pro hledání kořenů libovolné funkce, ale především je užitečná zejména pro určení kořenů polynomu.

Müllerova metoda je zobecněním metody sečen. Metoda sečen v podstatě znamená, že pro dané aproximace x^k, x^{k-1} bodu ξ aproximujeme funkci f přímkou procházející body $\left[x^{k-1}, f(x^{k-1})\right]$, $\left[x^k, f(x^k)\right]$ a za další aproximaci bodu ξ vezmeme průsečík této přímky s osou x. Müllerova metoda užívá tři aproximace x^{k-2}, x^{k-1}, x^k a křivku $y = f(x)$ aproximujeme parabolou určenou těmito body. Průsečík této paraboly s osou x, který je nejbližší k x^k, vezmeme za další aproximaci x^{k+1}. Touto metodou lze najít i násobné a komplexní kořeny. Lze ji popsat následujícím způsobem.

Nechť x^{k-2}, x^{k-1}, x^k jsou již vypočtené aproximace. Sestrojme polynom

$$P(x) = a(x - x^k)^2 + b(x - x^k) + c$$

procházející body $\left[x^{k-2}, f(x^{k-2})\right]$, $\left[x^{k-1}, f(x^{k-1})\right]$, $\left[x^k, f(x^k)\right]$, t.j. splňující podmínky $P(x^i) = f(x^i)$, $i = k-2, k-1, k$. Z nich plyne

$$c = f(x^k)$$

$$b = \frac{(x^{k-2} - x^k)^2 \left[f(x^{k-1}) - f(x^k)\right] - (x^{k-1} - x^k)^2 \left[f(x^{k-2}) - f(x^k)\right]}{(x^{k-2} - x^k)(x^{k-1} - x^k)(x^{k-2} - x^{k-1})}$$

$$a = \frac{(x^{k-2} - x^k) \left[f(x^{k-1}) - f(x^k)\right] - (x^{k-1} - x^k) \left[f(x^{k-2}) - f(x^k)\right]}{(x^{k-2} - x^k)(x^{k-1} - x^k)(x^{k-1} - x^{k-2})}$$

Kořeny kvadratické rovnice $P(x) = 0$ je vhodné vyjádřit (vzhledem k zaokrouhlovacím chybám) ve tvaru

$$x^{k+1} - x^k = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}}.$$

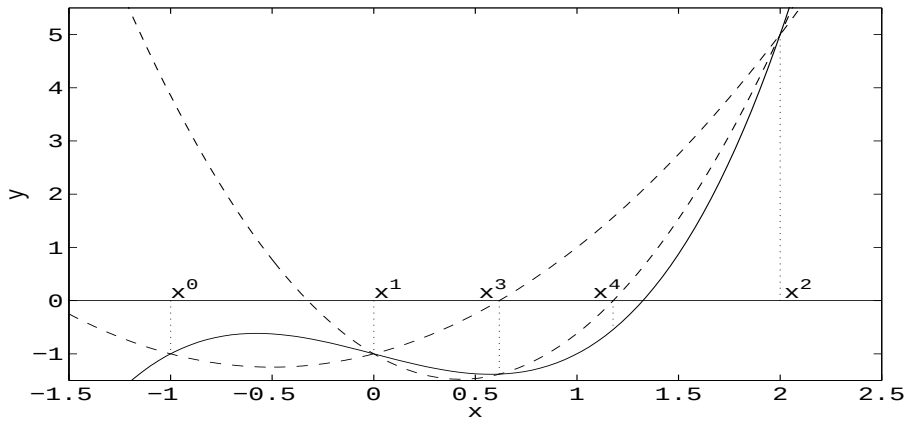
Znaménko u odmocniny vybereme tak, aby bylo shodné se znaménkem b. Tato volba znamená, že jmenovatel zlomku bude v absolutní hodnotě největší a tedy výsledná hodnota x^{k+1} bude nejbližší x^k. Je tedy

$$x^{k+1} = x^k - \frac{2c}{b + (\operatorname{sign} b)\sqrt{b^2 - 4ac}}.$$

Další postup opakujeme s aproximacemi x^{k-1}, x^k, x^{k+1} atd. Při řešení kvadratické rovnice musíme užívat komplexní aritmetiky, neboť výraz $b^2 - 4ac$ může být záporný.

Příklad 2.14. Aplikujme Müllerovu metodu na řešení rovnice $x^3 - x - 1 = 0$. Výpočet je uveden v tabulce a graficky zobrazen na obr. 2.21. Funkce $f(x) = x^3 - x - 1$ má jediný reálný kořen a proto lze v tomto případě vzít za počáteční aproximace x^0, x^1, x^2 (z důvodu vhodnějšího grafického vyjádření) body $x^0 = -1$, $x^1 = 0$, $x^2 = 2$.

k	x^k	$f(x^k)$
0	−1,00000000	−1,00000000
1	0,00000000	−1,00000000
2	2,00000000	5,00000000
3	0,61803399	−1,38196601
4	1,17827569	−0,54243597
5	1,30978731	−0,06279113
6	1,32509032	0,00158855
7	1,32471777	−0,00000081



Obr. 2.21: Müllerova metoda, $f(x) = x^3 - x - 1$

Lze ukázat [5], že Müllerova metoda je řádu alespoň $q = 1{,}84\ldots$, kde q je největší kořen rovnice $q^3 - q^2 - q - 1 = 0$.

§ 2.12. Iterační metody pro systémy nelineárních rovnic

Zabývejme se nyní řešením systému nelineárních rovnic. Je dáno m nelineárních

rovnic o m neznámých

$$\begin{aligned} f_1(x_1,\dots,x_m) &= 0 \\ &\vdots \\ f_m(x_1,\dots,x_m) &= 0 \end{aligned} \tag{2.40}$$

Kořenem systému (2.40) rozumíme každou uspořádanou m-tici reálných čísel $(\xi_1,\dots,\xi_m)$, která tomuto systému vyhovuje. Systém (2.40) lze také zapsat ve vektorovém tvaru

$$F(\boldsymbol{x}) = \boldsymbol{o}, \qquad \boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{o} = (0,\dots,0)^T \in \mathbb{R}^m. \tag{2.41}$$

Kořen této rovnice budeme nyní značit $\boldsymbol{\xi} = (\xi_1,\dots,\xi_m)^T$. Dále budeme postupovat obdobně jako při řešení jedné rovnice, tzn. rovnici (2.41) převedeme na rovnici ekvivalentní

$$\boldsymbol{x} = G(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^m \tag{2.42}$$

a budeme hledat pevný bod zobrazení $G : \mathbb{R}^m \to \mathbb{R}^m$. Systém (2.42) lze samozřejmě rozepsat takto:

$$\begin{aligned} x_1 &= g_1(x_1,\dots,x_m) \\ &\vdots \\ x_m &= g_m(x_1,\dots,x_m) \end{aligned} \tag{2.43}$$

Definujme nyní v prostoru $\mathbb{R}^m$ metriku vztahem

$$\varrho(\boldsymbol{x},\boldsymbol{y}) = \max_{1\le i\ \le m} |x_i - y_i|. \tag{2.44}$$

Prostor $\mathbb{R}^m$ s takto definovanou metrikou je úplným metrickým prostorem. Nyní lze pro vyšetřování konvergence iteračního procesu $\boldsymbol{x}^{k+1} = G(\boldsymbol{x}^k), k = 0,1,2,\dots$ užít Banachovy věty o pevném bodě.

Věta 2.18. *Nechť zobrazení $G : \mathbb{R}^m \to \mathbb{R}^m$ je kontrakce na $\mathbb{R}^m$,*

$$\varrho(G(\boldsymbol{x}),G(\boldsymbol{y})) \le q\varrho(\boldsymbol{x},\boldsymbol{y}), \qquad 0 \le q < 1.$$

Pak pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^m$ je posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$, $k = 0,1,2,\dots$, $\boldsymbol{x}^k = G(\boldsymbol{x}^{k-1})$, konvergentní v $\mathbb{R}^m$ a $\lim_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{\xi}$, kde $\boldsymbol{\xi}$ je jediný pevný bod zobrazení G.

Důkaz plyne ihned aplikací Banachovy věty o pevném bodě.

Užitečnější výsledek získáme, budeme-li předpokládat existenci kořene $\boldsymbol{\xi}$. Dokážeme tuto větu:

Věta 2.19. *Nechť $\boldsymbol{\xi} \in \mathbb{R}^m$ je pevný bod rovnice $\boldsymbol{x} = G(\boldsymbol{x})$. Nechť funkce $g_i, i = 1,\dots,m$, mají spojité parciální derivace pro všechna $\boldsymbol{x} \in \Omega(\boldsymbol{\xi},r)$, $\Omega(\boldsymbol{\xi},r) = \{\boldsymbol{x} | \varrho(\boldsymbol{x},\boldsymbol{\xi}) \le r\}$. Nechť dále platí*

$$\left|\frac{\partial g_i(\boldsymbol{x})}{\partial x_j}\right| \le \frac{q}{m}, \qquad i,j = 1,\dots,m, \tag{2.45}$$

$0 \le q < 1$ a nechť $\boldsymbol{x}^0 \in \Omega(\boldsymbol{\xi}, r)$. Pak všechny iterace $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ určené vztahem $\boldsymbol{x}^{k+1} = G(\boldsymbol{x}^k)$ leží v množině $\Omega(\boldsymbol{\xi}, r)$ a $\lim_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{\xi}$.

Důkaz. Množina $\Omega(\boldsymbol{\xi}, r)$ je uzavřenou omezenou podmnožinou úplného metrického prostoru $\mathbb{R}^m$ a je tedy rovněž úplným metrickým prostorem. Nyní je třeba ověřit, zda jsou na množině $\Omega(\boldsymbol{\xi}, r)$ splněny předpoklady Banachovy věty.

Pro libovolné dva body $\boldsymbol{x}, \boldsymbol{y} \in \Omega(\boldsymbol{\xi}, r)$ platí podle Taylorovy věty pro funkce více proměnných:

$$g_i(\boldsymbol{x}) - g_i(\boldsymbol{y}) = \sum_{j=1}^{m} \frac{\partial g_i(\boldsymbol{\alpha}_i)}{\partial x_j}(x_j - y_j),\ i = 1, \dots, m,$$

kde $\boldsymbol{\alpha}_i \in \Omega(\boldsymbol{\xi}, r)$, neboť $\boldsymbol{\alpha}_i$ leží na úsečce spojující body $\boldsymbol{x}$ a $\boldsymbol{y}$ a množina $\Omega(\boldsymbol{\xi}, r)$ je konvexní.

Pomocí vztahu (2.45) odhadneme nyní absolutní hodnotu rozdílu $g_i(\boldsymbol{x}) - g_i(\boldsymbol{y})$:

$$|g_i(\boldsymbol{x}) - g_i(\boldsymbol{y})| \le \sum_{j=1}^{m} \left| \frac{\partial g_i(\boldsymbol{\alpha}_i)}{\partial x_j} \right| |x_j - y_j| \le$$

$$\le \max_{1\le j\le m} |x_j - y_j| \sum_{j=1}^{m} \left| \frac{\partial g_i(\boldsymbol{\alpha}_i)}{\partial x_j} \right| \le q\varrho(\boldsymbol{x}, \boldsymbol{y}), \qquad i = 1, \dots, m.$$

Jelikož tato nerovnost platí pro všechna $i = 1, \dots, m$, je také

$$\max_{1\le i\ \le m} |g_i(\boldsymbol{x}) - g_i(\boldsymbol{y})| = \varrho(G(\boldsymbol{x}), G(\boldsymbol{y})) \le q\varrho(\boldsymbol{x}, \boldsymbol{y}).$$

Zobrazení G je tedy kontrakce na $\Omega(\boldsymbol{\xi}, r)$.

Nechť nyní $\boldsymbol{x} \in \Omega(\boldsymbol{\xi}, r)$, pak také $G(\boldsymbol{x}) \in \Omega(\boldsymbol{\xi}, r)$, neboť

$$\varrho(G(\boldsymbol{x}), \boldsymbol{\xi}) = \varrho(G(\boldsymbol{x}), G(\boldsymbol{\xi})) \le q\varrho(\boldsymbol{x}, \boldsymbol{\xi}) < r.$$

Odtud plyne, že pro zobrazení G platí $G : \Omega(\boldsymbol{\xi}, r) \to \Omega(\boldsymbol{\xi}, r)$. Tedy zobrazení G splňuje na úplném metrickém prostoru $\Omega(\boldsymbol{\xi}, r)$ předpoklady věty 2.18 a odtud plyne tvrzení věty. □

Poznámka 13. Je zřejmé, že předpoklad (2.45) lze nahradit předpokladem

$$\max_{1\le i\ \le m} \sum_{j=1}^{m} \left| \frac{\partial g_i(\boldsymbol{x})}{\partial x_j} \right| \le q < 1. \tag{2.46}$$

Příklad 2.15. Řešte systém nelineárních rovnic

$$\begin{aligned} x_1^2 - 2x_1 - x_2 + 0,5 &= 0 \\ x_1^2 + 4x_2^2 - 4 &= 0 \end{aligned}$$

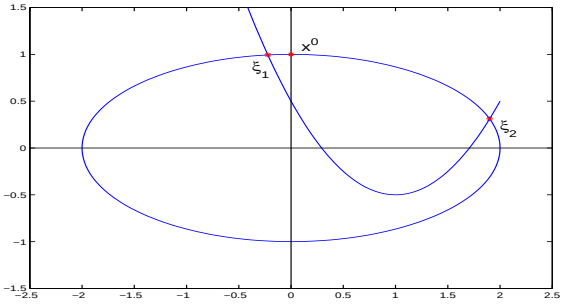
Řešení: První rovnici lze upravit následujícím způsobem:

$$\begin{aligned} x_1^2 - 2x_1 - x_2 + 0,5 &= 0 \\ x_2 &= x_1^2 - 2x_1 + 0,5 \\ x_2 &= (x_1 - 1)^2 - 0,5\,, \end{aligned}$$

jedná se tedy o rovnici paraboly. Druhá rovnice je rovnice elipsy

$$\frac{x_1^2}{4} + x_2^2 = 1\,.$$

Hledáme tedy průsečíky paraboly s elipsou. Průsečík $\boldsymbol{\xi}_1$ leží ve druhém kvadranru, průsečík $\boldsymbol{\xi}_2$ v prvním kvadrantu (viz obrázek 2.22). Budeme hledat aproximaci průsečíku $\boldsymbol{\xi}_1$. Rovnice převedeme na tvar



Obr. 2.22: Průsečíky elipsy $\frac{x_1^2}{4} + x_2^2 = 1$ s parabolou $x_2 = (x_1 - 1)^2 - 0,5$

$$\begin{aligned} x_1 &= g_1(x_1, x_2) = \frac{x_1^2 - x_2 + 0,5}{2} \\ x_2 &= g_2(x_1, x_2) = \frac{-x_1^2 - 4x_2^2 + 8x_2 + 4}{8} \end{aligned}$$

Ve druhé rovnici byl na každou stranu přidán člen $-8x_2$ a rovnice pak vydělena -8. Jedná se o poněkud umělý krok, který ovšem zajistí dostatečné podmínky pro konvergenci.
Jako počáteční iteraci zvolíme bod $(x_1^0, x_2^0) = (0; 1)$, který leží blízko $\boldsymbol{\xi}_1$.V okolí hledaného průsečíku jsou zřejmě splněny podmínky (2.45). Postupně dostaneme

$$\begin{aligned} x_1^1 &= g_1(x_1^0, x_2^0) = -0,25, & x_2^1 &= g_2(x_1^0, x_2^0) = 1 \\ x_1^2 &= g_1(x_1^1, x_2^1) = -0,21875, & x_2^2 &= g_2(x_1^1, x_2^1) = 0,9921875 \\ &\vdots \\ x_1^8 &= g_1(x_1^7, x_2^7) \doteq -0,2222145, & x_2^8 &= g_2(x_1^7, x_2^7) \doteq 0,9938084 \\ x_1^9 &= g_1(x_1^8, x_2^8) \doteq -0,2222146, & x_2^9 &= g_2(x_1^8, x_2^8) \doteq 0,9938084 \end{aligned}$$

Pokud bychom chtěli najít druhý průsečík, museli bychom změnit iterační funkce, abychom zajistili konvergenci. Možná volba je např.

$$\begin{aligned} x_1 &= g_1(x_1, x_2) = \frac{-x_1^2 + 4x_1 + x_2 - 0,5}{2} \\ x_2 &= g_2(x_1, x_2) = \frac{-x_1^2 - 4x_2^2 + 11x_2 + 4}{11}\,. \end{aligned}$$

Navrhneme nyní modifikaci iterační metody $\boldsymbol{x}^{k+1} = G(\boldsymbol{x}^k)$. Tato modifikace spočívá v tom, že pro výpočet x_i^{k+1} použijeme již vypočtených hodnot $x_1^{k+1}, \ldots, x_{i-1}^{k+1}$, tj.

$$\begin{aligned} x_1^{k+1} &= g_1(x_1^k, x_2^k, \ldots, x_m^k) \\ x_2^{k+1} &= g_2(x_1^{k+1}, x_2^k, \ldots, x_m^k) \\ x_3^{k+1} &= g_3(x_1^{k+1}, x_2^{k+1}, \ldots, x_m^k) \\ &\vdots \\ x_m^{k+1} &= g_m(x_1^{k+1}, \ldots, x_{m-1}^{k+1}, x_m^k). \end{aligned}$$

Tato modifikace se nazývá *Seidelova metoda*.

Otázkami konvergence této metody se zabývá monografie [17].

§ 2.13. Newtonova metoda pro systémy nelineárních rovnic

Rovnici $F(\boldsymbol{x}) = \boldsymbol{o}$ lze převést na rovnici $\boldsymbol{x} = G(\boldsymbol{x})$ různými způsoby. Pojednáme podrobněji o tvaru *Newtonovy metody* pro systémy nelineárních rovnic.

Položme

$$G(\boldsymbol{x}) = \boldsymbol{x} - J_F^{-1}(\boldsymbol{x})F(\boldsymbol{x}),$$

kde

$$J_F(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \dfrac{\partial f_m(\boldsymbol{x})}{\partial x_m} \end{pmatrix}$$

Matici $J_F(\boldsymbol{x})$ nazýváme *Jacobiovou maticí* funkce F. Nechť $J_F(\boldsymbol{x})$ je regulární matice se spojitými prvky v okolí bodu $\boldsymbol{\xi}$. Iterační metodu

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - J_F^{-1}(\boldsymbol{x}^k)F(\boldsymbol{x}^k), \qquad k = 0, 1, 2, \ldots \tag{2.47}$$

nazýváme *Newtonovou iterační metodou* pro systém $F(\boldsymbol{x}) = \boldsymbol{o}$.

Otázky konvergence a volby počáteční aproximace pro metodu (2.47) jsou mnohem komplikovanější než v případě jedné rovnice. Tato problematika je podrobně studována v [5] nebo v [17]. Zde uvedeme bez důkazu jednu ze základních vět.

Věta 2.20. *Nechť $\boldsymbol{\xi}$ je kořenem rovnice $F(\boldsymbol{x}) = \boldsymbol{o}$. Nechť $J_F(\boldsymbol{x})$ je regulární matice se spojitými prvky v okolí $O(\boldsymbol{\xi})$ bodu $\boldsymbol{\xi}$, přičemž*

$$\left\|J_F^{-1}(\boldsymbol{x})\right\|_\infty \leq K, \qquad K = \text{konst.},$$

pro všechna $\boldsymbol{x}$ z tohoto okolí. Nechť funkce $f_i, i = 1, \ldots, m$, mají spojité druhé parciální derivace v $O(\boldsymbol{\xi})$.

Posloupnost $\left\{\boldsymbol{x}^k\right\}_{k=0}^{\infty}$ určená Newtonovou metodou konverguje ke kořenu $\boldsymbol{\xi}$ za předpokladu, že počáteční aproximace $\boldsymbol{x}^0$ leží dostatečně blízko $\boldsymbol{\xi}$. Řád metody je roven dvěma.

Důkaz viz [2], [5].

Tato věta předpokládá existenci „dobré“ počáteční aproximace. Ovšem pro systém nelineárních rovnic není snadné určit „dobrou“ počáteční aproximaci. Pro systém $F(\boldsymbol{x}) = \boldsymbol{o}$ neexistují jednoduché, vždy konvergentní metody, které by umožnily vhodnou volbu počáteční aproximace. Lze sice formulovat podmínky pro počáteční aproximaci, pro kterou bude iterační proces konvergovat, ale ověření těchto podmínek je značně náročné a pro praktické výpočty nepoužitelné.

V případě dvou rovnic lze v některých případech určit počáteční aproximaci geometricky.

Poznámka 14. Problém řešení systému nelineárních rovnic lze převést na problém minimalizace funkce více proměnných, viz např. [5], [17].

Vraťme se nyní ke vztahu (2.47). Zde je třeba na každém kroku vypočítat inverzní matici $J_F^{-1}(\boldsymbol{x})$. Tento vztah však můžeme zapsat i takto:

$$J_F(\boldsymbol{x}^k)(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) = -F(\boldsymbol{x}^k). \tag{2.48}$$

Položíme-li nyní $\boldsymbol{\delta}^k = \boldsymbol{x}^{k+1} - \boldsymbol{x}^k$, tj. $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{\delta}^k$, $k = 0, 1, 2, \ldots$, dostaneme

$$J_F(\boldsymbol{x}^k)\boldsymbol{\delta}^k = -F(\boldsymbol{x}^k); \qquad \boldsymbol{\delta}^k = (\delta_1^k, \ldots, \delta_m^k)^T. \tag{2.49}$$

(2.49) je nyní systém lineárních rovnic pro neznámé $\delta_1^k, \ldots, \delta_m^k$. Řešení tohoto systému, vektor $\boldsymbol{\delta}^k$ je vektor „oprav“ pro aproximaci $\boldsymbol{x}^k$.

Příklad 2.16. Použijte Newtonovu metodu k nalezení průsečíku $\boldsymbol{\xi}_2$ paraboly a elipsy z příkladu 2.15
Řešení: Položme

$$F(x_1, x_2) = \begin{pmatrix} x_1^2 - 2x_1 - x_2 + 0,5 \\ x_1^2 + 4x_2^2 - 4 \end{pmatrix}, \quad \text{pak } J_F(x_1, x_2) = \begin{pmatrix} 2x_1 - 2 & -1 \\ 2x_1 & 8x_2 \end{pmatrix}.$$

Zvolme počáteční aproximaci $(x_1^0, x_2^0) = (2; 0,25)$, takže

$$F(2; 0,25) = \begin{pmatrix} 0,25 \\ 0,25 \end{pmatrix}, \quad J_F(2; 0,25) = \begin{pmatrix} 2,0 & -1,0 \\ 4,0 & 2,0 \end{pmatrix}.$$

Řešíme tedy systém

$$\begin{pmatrix} 2,0 & -1,0 \\ 4,0 & 2,0 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} = - \begin{pmatrix} 0,25 \\ 0,25 \end{pmatrix}$$

Řešení je $\delta_1 = -0.09375$, $\delta_2 = 0.0625$, tj.

$$\boldsymbol{x}^1 = \begin{pmatrix} 2,0 - 0.09375 \\ 0,25 + 0.0625 \end{pmatrix} = \begin{pmatrix} 1.90625 \\ 0,3125 \end{pmatrix}.$$

Dalším postupem dostaneme

$$\boldsymbol{x}^2 \doteq \begin{pmatrix} 1.900691 \\ 0.311213 \end{pmatrix}, \quad \boldsymbol{x}^3 \doteq \begin{pmatrix} 1.900677 \\ 0.311219 \end{pmatrix}, \quad \boldsymbol{x}^4 \doteq \begin{pmatrix} 1.900677 \\ 0.311219 \end{pmatrix}.$$

Poznámka 15. Jako kritérium pro dosažení dostatečně malé chyby aproximace a zastavení výpočtu můžeme použít např. odhad $\frac{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|}{\|\boldsymbol{x}^k\|} \leq \varepsilon$, kde ε je předem daná přesnost výpočtu.

Cvičení ke kapitole 2

1. Funkce $f(x) = x^3 + 4x^2 - 10$ má jediný kořen v intervalu $[1; 1{,}5]$. Uvažujte tyto iterační funkce pro nalezení kořene ($\xi \approx 1{,}365230013$):

$$g_1(x) = x - x^3 - 4x^2 + 10 \qquad g_4(x) = \left(\frac{10}{4+x}\right)^{\frac{1}{2}}$$
$$g_2(x) = \left(\frac{10}{x} - 4x\right)^{\frac{1}{2}} \qquad g_5(x) = x - \frac{x^3 + 4x^2 - 10}{3x^2 + 8x}$$
$$g_3(x) = \frac{1}{2}(10 - x^3)^{\frac{1}{2}}$$

   Nechť počáteční aproximace $x_0 = 1{,}5$. Ukažte, že funkce g_3, g_4, g_5 jsou vhodné iterační funkce (tj. posloupnost iterací konverguje ke kořenu ξ). Dále ukažte, že volba funkce g_1 vede na divergentní posloupnost a posloupnost $\{x^k\}$, $x^k = g_2(x^{k-1})$, $x^0 = 1{,}5$, není definována (v oboru reálných čísel).

2. Ukažte, že funkce $g(x) = 2^{-x}$ má jediný pevný bod v intervalu $[\frac{1}{3}; 1]$. Najděte tento pevný bod s chybou menší než 10^{-4}. Kolik iterací je třeba k dosažení této přesnosti?
   (Řešení: $x^0 = 1$, $x^{12} = 0{,}6412053$. Podle důsledku věty 2.5 je třeba 15 iterací.)

3. Je dána rovnice $3x^2 - e^x = 0$. Určete interval, ve kterém leží kladný kořen této rovnice. Najděte vhodnou iterační funkci g, pro kterou iterační metoda $x^{k+1} = g(x^k)$ bude konvergovat k tomuto kladnému kořenu.

4. Je dána rovnice $3x^3 - x - 1 = 0$. Určete interval, ve kterém leží kladný kořen této rovnice. Najděte vhodnou iterační funkci g, pro kterou iterační metoda $x^{k+1} = g(x^k)$ bude konvergovat k tomuto kladnému kořenu.

5. Je dána rovnice $x = g(x)$, $g(x) = (6 + x)^{1/2}$. Pevný bod je $\xi = (3, 3)$. Znázorněte geometricky příslušný iterační proces $x^{k+1} = g(x^k)$, $x^0 = 7$. Bude tento iterační proces konvergovat?

6. Ukažte, že iterační funkce

$$g(x) = \frac{1}{2}\left(x + \frac{2}{x}\right)$$

splňuje podmínky pro výpočet $\sqrt{2}$. Jaký tvar má funkce g pro výpočet $\sqrt{a}$, $a > 0$?

7. Užitím Newtonovy metody vypočtěte $\sqrt{13}$. Zvolte vhodnou funkci a počáteční aproximaci.

8. Newtonovou metodou nalezněte kořen funkce $f(x) = x - \cos x$.
(Řešení: $x^0 = \frac{\pi}{4}$, $x^3 = 0{,}7390851332$.)

9. Užitím Newtonovy metody s počáteční aproximací $x^0 = 10$ vypočtěte $\sqrt{91}$.
(Řešení: $x^{k+1} = (x^k + 91/x^k)/2$, $x^3 = 9{,}539392015$.)

10. Na parabole $y = x^2$ najděte užitím Newtonovy metody bod nejbližší bodu $(1, 3)$.
Návod:
    1. Určete druhou mocninu vzdálenosti $d^2(x)$ bodu $X = (x, x^2)$ ležícího na parabole a bodu $(1, 3)$.
    2. Řešte rovnici $(d^2(x))' = f(x) = 0$. Za počáteční aproximaci zvolte $x^0 = 1, 0$.
    ($x^4 \doteq 1, 28962390$)

11. Užijte a) Newtonovy metody, b) metody sečen, c) metody regula falsi k nalezení kořenů funkcí

    1) $x^3 - 2x^2 - 5 = 0, \quad \xi \in [1, 4]$,
    2) $x - 0{,}8 - 0{,}2 \sin x = 0, \quad \xi \in [0, \frac{\pi}{2}]$,
    3) $3x^2 - \mathrm{e}^x = 0, \quad \xi \in [0, 2]$.

12. Je dána rovnice $\frac{4x-7}{x-2} = 0$.

    (a) Jaké jsou vhodné počáteční aproximace pro metodu regula falsi?
    (b) Je $x^0 = 3$ vhodná počáteční aproximace pro použití Newtonovy metody?

13. Metodou regula falsi najděte kladný kořen rovnice $x^2 - 7 = 0$.

14. Řešte rovnici $x = 2^{-x}$ pro kořen v intervalu $[0, 1]$ Steffensenovou metodou a porovnejte s výsledkem cvičení 2.

15. Vypočtěte $\sqrt{3}$ s počáteční aproximací $x^0 = 2$ Steffensenovou metodou a porovnejte výsledky s použitím metody Newtonovy, metody sečen a metody regula falsi.

16. Užijte Müllerovy metody k nalezení kořenů rovnice $x^3 + 3x^2 - 1 = 0$.
(Řešení: 0,532089, $-0{,}652703$, $-2{,}87938$.)

17. Je dán systém nelineárních rovnic

$$x_1^2 - x_2 - 0{,}2 = 0,$$
$$x_2^2 - x_1 - 0{,}3 = 0.$$

Užitím Newtonovy metody nalezněte kořen ležící v 1. kvadrantu. Počáteční aproximaci určete graficky.
($\boldsymbol{x}^0 = (1{,}2; 1{,}2)^T$, $\boldsymbol{x}_2 = (1{,}192309; 1{,}221601)^T$.)

18. Uvažujme systém nelineárních rovnic

$$x_1 = \frac{2x_1 - x_1^2 + x_2}{2} \quad \text{(parabola)},$$
$$x_2 = \frac{2x_1 - x_1^2 + 8}{9} + \frac{4x_2 - x_2^2}{4} \quad \text{(elipsa)}.$$

Zvolte $\boldsymbol{x}^0 = (1{,}4; 2{,}0)^T$ a vypočtěte 2 iterace

a) iterační metodou $\boldsymbol{x}^k = G(\boldsymbol{x}^{k-1})$,

b) Seidelovou metodou.

Výsledky porovnejte s přesným řešením $\boldsymbol{\xi} = (1{,}4076401; 1{,}9814506)^T$.

19. Je dána funkce soustava nelineárních rovnic

$$\begin{aligned} x_1 &= \frac{7x_1^3 - x_2 - 1}{10} \equiv g_1(x_1, x_2) \\ x_2 &= \frac{8x_2^3 + x_1 - 1}{11} \equiv g_2(x_1, x_2) \end{aligned}$$

Tato soustava má 9 pevných bodů.
Ověřte, že v okolí bodu (0,0) splňuje tato soustava podmínku pro konvergenci iteračního procesu

$$\begin{aligned} x_1^{k+1} &= g_1(x_1^k, x_2^k) \\ x_2^{k+1} &= g_2(x_1^k, x_2^k) \end{aligned}.$$

Bude tato podmínka splněna v okolí bodu (1,1)?

Kontrolní otázky ke kapitole 2

1. Je možné použít prostou iterační metodu v případě, že funkce g zobrazuje interval $I = [a, b]$ do sebe a platí $|g'(x)| \leq 1$, přičemž rovnost nastává pouze v některém z krajních bodů intervalu I? Proč?

2. Je dána funkce $f(x) = \cos x$.
   Newtonovou metodou chceme najít kořen $\xi = \frac{3\pi}{2}$. Můžeme použít počáteční aproximaci $x^0 = 3$? Proč?
   Můžeme použít počáteční aproximaci $x^0 = 5$? Proč?

3. Je dána funkce $f(x) = (x - 3)^{1/2}$.
   Můžeme užít Newtonovu metodu pro nalezení kořene s počáteční aproximací $x^0 = 4$? Proč?

4. Co se stane, když použijeme Müllerovu metodu pro nalezení kořene polynomu 2. stupně?

5. Co se stane při použití Newtonovy metody pro systém nelineárních rovnic na řešení soustavy rovnic lineárních?

6. Aplikujte Aitkenovou δ^2-metodou na posloupnost částečných součtů geometrické řady. Co nastane a proč?

Kapitola 3

Polynomy

Nechť Π_n je třída polynomů stupně nejvýše n s reálnými koeficienty. Polynom $P \in \Pi_n$ budeme zapisovat ve tvaru

$$P(x) = a_0 x^n + \ldots + a_n. \tag{3.1}$$

Označme $\xi_1, \xi_2, \ldots, \xi_n$ kořeny (reálné i komplexní) polynomu P. V této kapitole se budeme zabývat numerickými metodami pro určení těchto kořenů.

§ 3.1. Hranice kořenů

Pro volbu počáteční aproximace pro aplikaci numerické metody je vhodné znát hranice těchto kořenů:

Věta 3.1. *Nechť*

$$A = \max\left(|a_1|, \ldots, |a_n|\right),$$
$$B = \max\left(|a_0|, \ldots, |a_{n-1}|\right),$$

kde a_k, $k = 0, 1, \ldots, n$, $a_0 a_n \neq 0$, jsou koeficienty polynomu $P \in \Pi_n$, Pak pro všechny kořeny $\xi_k, k = 0, 1, \ldots, n$, polynomu P platí

$$\frac{1}{1 + \dfrac{B}{|a_n|}} \leq |\xi_k| \leq 1 + \frac{A}{|a_0|}. \tag{3.2}$$

Důkaz. Bez újmy na obecnosti předpokládejme $|x| > 1$. Pak

$$\begin{aligned} |P(x)| &\geq |a_0 x^n| - \left(\left|a_1 x^{n-1}\right| + \ldots + |a_n|\right) \geq \\ &\geq |a_0||x|^n - A\left(|x|^{n-1} + \ldots + 1\right) = \\ &= |a_0||x|^n - A\frac{|x|^n - 1}{|x| - 1} > \left(|a_0| - \frac{A}{|x| - 1}\right)|x|^n. \end{aligned}$$

Jestliže $|a_0|-A/(|x|-1)>0$, pak je $|P(x)|>0$ a to znamená, že pro $|x|>1+A/|a_0|$ nemá P kořen. Odtud plyne, že všechny kořeny polynomu P splňují nerovnost

$$|\xi_k| \le 1+\frac{A}{|a_0|}, \qquad k=1,\dots,n.$$

Položme nyní $x=1/y$. Užitím této substituce dostaneme:

$$P(x)=P(\frac{1}{y})=\frac{1}{y^n}\left(a_0+\ldots+a_n y^n\right)=\frac{1}{y^n}Q(y).$$

Určíme horní hranici kořenů $\eta_1,\dots,\eta_n$ polynomu $Q\in\Pi_n$. Podle předchozího je

$$|\eta_k| \le 1+\frac{B}{|a_n|}, \qquad k=1,\dots,n.$$

Jelikož $\eta_k=1/\xi_k$, $k=1,\dots,n$, plyne odtud požadovaná nerovnost (3.2). □

Příklad 3.1. Určete hranice reálných kořenů polynomu

$$P(x)=x^6-2x^5+8x^4+3x^3-x^2+x-10.$$

Řešení.

$$A=\max(2,8,3,1,1,10)=10$$
$$B=\max(1,2,8,3,1,1)=8$$

Odtud

$$\frac{5}{9}=\frac{1}{1+\frac{8}{10}}\le|\xi_k|\le 1+\frac{10}{1}=11.$$

Některé další hranice pro kořeny polynomu jsou uvedeny v následující větě:

Věta 3.2. *Pro všechny kořeny $\xi_k, k=0,1,\dots,n$, polynomu $P\in\Pi_n$ platí*

$$|\xi_k| \le \max\left\{1,\sum_{j=1}^{n}\left|\frac{a_j}{a_0}\right|\right\}$$
$$|\xi_k| \le 2\max\left\{\left|\frac{a_1}{a_0}\right|,\sqrt{\left|\frac{a_2}{a_0}\right|},\sqrt[3]{\left|\frac{a_3}{a_0}\right|},\dots,\sqrt[n]{\left|\frac{a_n}{a_0}\right|}\right\}$$
$$|\xi_k| \le \max\left\{\left|\frac{a_n}{a_0}\right|,1+\left|\frac{a_{n-1}}{a_0}\right|,\dots,1+\left|\frac{a_1}{a_0}\right|\right\}.$$

Důkaz lze najít např. v [5].

§ 3.2. Počet reálných kořenů polynomu

Zabývejme se nyní otázkou počtu reálných kořenů polynomu.

Úmluva. Nechť $c_1, \ldots, c_m$ je posloupnost reálných čísel různých od nuly.

Řekneme, že pro dvojici c_k, c_{k+1} *nastává znaménková změna*, jestliže $c_k c_{k+1} < 0$.

Řekneme, že dvojice c_k, c_{k+1} *zachovává znaménko*, jestliže $c_k c_{k+1} > 0$.

Dále ukážeme, že počet reálných kořenů polynomu lze určit pomocí posloupnosti polynomů klesajících stupňů. Vhodnou posloupností je tzv. Sturmova posloupnost:

Definice 3.1. Posloupnost reálných polynomů

$$P = P_0, P_1, \ldots, P_m$$

se nazývá *Sturmovou posloupností* příslušnou polynomu P, jestliže

(a) Všechny reálné kořeny polynomu P_0 jsou jednoduché.

(b) Je-li ξ reálný kořen polynomu P_0, pak $\operatorname{sign} P_1(\xi) = -\operatorname{sign} P_0'(\xi)$.

(c) Pro $i = 1, 2, \ldots, m-1$,

$$P_{i+1}(\alpha) P_{i-1}(\alpha) < 0,$$

jestliže α je reálný kořen polynomu P_i.

(d) Poslední polynom P_m nemá reálné kořeny.

Poznámka 1. Jestliže polynom P má násobné kořeny, pak dělením polynomu P největším společným dělitelem P a P' dostaneme polynom, který má tytéž kořeny, ale všechny jednoduché.

Uveďme nyní jednoduchý postup pro konstrukci Sturmovy posloupnosti příslušné polynomu P za předpokladu, že všechny reálné kořeny polynomu P jsou jednoduché.

Položme

$$P_0(x) = P(x), \qquad P_1(x) = -P_0'(x) \tag{3.3}$$

a sestrojme další polynomy P_{i+1} rekurentně dělením polynomu P_{i-1} polynomem P_i:

$$P_{i-1}(x) = Q_i(x) P_i(x) - c_i P_{i+1}(x), \qquad i = 1, 2, \ldots, \tag{3.4}$$

kde

$$\text{stupeň } P_i > \text{stupeň } P_{i+1}$$

a konstanty c_i jsou kladné, ale jinak libovolné. Lze říci, že P_{i+1} je záporně vzatý zbytek při dělení P_{i-1}/P_i. Tato rekurze je známý Euklidův algoritmus. Protože stupně polynomů klesají, musí algoritmus končit po $m \leq n$ krocích:

$$P_{m-1}(x) = Q_m(x) P_m(x), \qquad P_m(x) \not\equiv 0.$$

Poslední polynom P_m je největší společný dělitel polynomů P a $P_1 = -P'$. Jestliže všechny reálné kořeny polynomu P jsou jednoduché, pak P a P' nemají žádné společné reálné kořeny. Tedy P_m nemá reálné kořeny.

Jestliže $P_i(\alpha) = 0$, tak z (3.4) plyne

$$P_{i-1}(\alpha) = -c_i P_{i+1}(\alpha).$$

Jestliže bychom předpokládali, že $P_{i+1}(\alpha) = 0$, pak z (3.4) by plynulo $P_{i+1}(\alpha) = \ldots = P_m(\alpha) = 0$, což by byl spor, neboť $P_m(\alpha) \neq 0$. Odtud tedy plyne $P_{i-1}(\alpha)P_{i+1}(\alpha) < 0$, je-li $P_i(\alpha) = 0$. Výše zkonstruovaná posloupnost zřejmě také splňuje podmínku 2 z definice Sturmovy posloupnosti a je tedy Sturmovou posloupností příslušnou polynomu P.

Věta 3.3. (Sturm). *Počet reálných kořenů polynomu P v intervalu $a \leq x < b$ je roven $W(b) - W(a)$, kde $W(x)$ je počet znaménkových změn ve Sturmově posloupnosti $P_0(x), \ldots, P_m(x)$ v bodě x (z níž jsou vyškrtnuty nuly).*

Důkaz. Nejdříve ukážeme, jaký vliv má malá změna hodnoty a na počet znaménkových změn $W(a)$ v posloupnosti

$$P_0(a), P_1(a), \ldots, P_m(a). \tag{3.5}$$

Nechť nejdříve a není kořenem žádného z polynomů P_i, $i = 0, \ldots, m$. Pak samozřejmě malá změna čísla a nemá žádný vliv na počet znaménkových změn v posloupnosti (3.5).

Nechť nyní a je kořenem některého z polynomů P_i, $i = 0, 1, \ldots, m-1$. Předpokládejme nejdříve, že $1 \leq i \leq m-1$. Nechť polynom P_i mění znaménko v bodě a. Pro dostatečně malé $h > 0$ zachycují chování polynomů P_{i-1}, P_i, P_{i+1} v bodech $a-h, a, a+h$ následující tabulky.

	$a-h$	a	$a+h$
P_{i-1}	$-$	$-$	$-$
P_i	$-$	0	$+$
P_{i+1}	$+$	$+$	$+$
$W(x)$	1	1	1

	$a-h$	a	$a+h$
P_{i-1}	$+$	$+$	$+$
P_i	$-$	0	$+$
P_{i+1}	$-$	$-$	$-$
$W(x)$	1	1	1

	$a-h$	a	$a+h$
P_{i-1}	$-$	$-$	$-$
P_i	$+$	0	$-$
P_{i+1}	$+$	$+$	$+$
$W(x)$	1	1	1

	$a-h$	a	$a+h$
P_{i-1}	$+$	$+$	$+$
P_i	$+$	0	$-$
P_{i+1}	$-$	$-$	$-$
$W(x)$	1	1	1

Ve všech těchto případech $W(a-h) = W(a) = W(a+h)$, což znamená, že při přechodu přes bod a, který je kořenem některého z polynomů P_i, $i = 1, \ldots, m-1$,

počet znaménkových změn ve Sturmově posloupnosti se nemění. Totéž platí i v případě, že P_i nemění znaménko v bodě a.

Nechť nyní a je kořenem polynomu P_0. Znaménka polynomů P_0, P_1 v okolí bodu a ukazují následující tabulky.

	$a-h$	a	$a+h$
P_0	$-$	0	$+$
P_1	$-$	$-$	$-$
$W(x)$	0	0	1

	$a-h$	a	$a+h$
P_0	$+$	0	$-$
P_1	$+$	$+$	$+$
$W(x)$	0	0	1

V obou případech $W(a+h) - W(a-h) = 1$. Odtud plyne: Je-li a kořenem polynomu P_0, pak při přechodu přes bod a získáme jednu znaménkovou změnu.

Pro $a < b$ a dostatečně malé $h > 0$ číslo

$$W(b) - W(a) = W(b-h) - W(a-h)$$

udává počet kořenů polynomu P v intervalu $a-h < x < b-h$. Protože číslo $h > 0$ je libovolně malé, udává tento rozdíl rovněž počet kořenů v intervalu $a \le x < b$. □

Příklad 3.2. Určete počet reálných kořenů polynomu

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

Řešení. Sestrojíme Sturmovu posloupnost příslušnou polynomu $P(x)$. Je

$$P_0(x) = x^3 - 3x + 1, \qquad P_0'(x) = 3x^2 - 3,$$
$$P_1(x) = -x^2 + 1.$$

Polynom P_2 je záporně vzatý zbytek při dělení polynomu P_0 polynomem P_1, tj. $P_2(x) = 2x - 1$ a dále $P_3(x) = -3/4$.

Sestavíme tabulku pro určení počtu reálných kořenů.

x	$P_0(x)$	$P_1(x)$	$P_2(x)$	$P_3(x)$	$W(x)$
$-\infty$	$-$	$-$	$-$	$-$	0
$+\infty$	$+$	$-$	$+$	$-$	3
0	$+$	$+$	$-$	$-$	1
-1	$+$	0	$-$	$-$	1
-2	$-$	$-$	$-$	$-$	0
1	$-$	0	$+$	$-$	2
2	$+$	$-$	$+$	$-$	3

Odtud plyne: $W(\infty) - W(-\infty) = 3$ $\Rightarrow$ 3 reálné kořeny
$W(\infty) - W(0) = 2$ $\Rightarrow$ 2 kladné kořeny
$W(-1) - W(-2) = 1$ $\Rightarrow$ 1 kořen v intervalu $[-2, -1]$
$W(1) - W(0) = 1$ $\Rightarrow$ 1 kořen v intervalu $[0, 1]$
$W(2) - W(1) = 1$ $\Rightarrow$ 1 kořen v intervalu $[1, 2]$

I když z teoretického hlediska vypadá konstrukce Sturmovy posloupnosti velmi jednoduše, mohou být konkrétní výpočty poněkud těžkopádné, neboť koeficienty polynomů P_i mohou být řádově dosti velké. Pro rychlý odhad počtu kladných kořenů daného polynomu je vhodná následující Descartesova věta.

Věta 3.4. (Descartes). *Počet kladných kořenů polynomu P (počítáno s násobností) je roven počtu znaménkových změn v posloupnosti koeficientů $a_0, \ldots, a_n$ nebo o sudé číslo menší.*

Jsou-li všechny koeficienty $a_0, \ldots, a_n$ různé od nuly, pak počet záporných kořenů je roven počtu zachování znamének v této posloupnosti nebo o sudé číslo menší.

Důkaz viz [2].

Příklad 3.3. Odhadněte počet kladných a záporných kořenů polynomu

$$P(x) = x^6 - 2x^5 + 8x^4 + 3x^3 - x^2 + x - 10$$

Posloupnost koeficientů: $1, -2, 8, 3, -1, 1, -10$
Počet kladných kořenů: 5 nebo 3 nebo 1
Počet záporných kořenů: 1

§ 3.3. Newtonova metoda a její modifikace

Pro určení kořenů polynomu P lze použít kterékoliv z metod uvedených v předchozí kapitole. Pro polynomy jsou vhodné zejména metoda Newtonova a Müllerova. Připomeňme, že Newtonova metoda je tvaru

$$x^{k+1} = x^k - \frac{P(x^k)}{P'(x^k)}, \qquad k = 0, 1, \ldots$$

Hodnoty $P(x^k)$, $P'(x^k)$ lze snadno spočítat Hornerovým schematem.

Hornerovo schema.

Nechť α je reálné číslo. Vydělíme polynom $P(x)$ lineárním polynomem $x - \alpha$:

$$P(x) = (x - \alpha)Q(x) + b_n,$$

kde

$$Q(x) = b_0 x^{n-1} + \ldots + b_{n-1}\,.$$

Koeficienty b_i, $i = 0, \ldots, n$ určíme z rekurentních vztahů:

$$\begin{aligned} b_0 &= a_0 \\ b_k &= a_k + \alpha b_{k-1}, \quad k = 1, \ldots, n\,. \end{aligned}$$

Pak je zřejmě $P(\alpha) = b_n$.

Poznámka 2. Nechť $\bar{\xi}_1$ je přibližný kořen polynomu P. Polynom P dělíme lineárním polynomem $(x - \bar{\xi}_1)$, tj.

$$P(x) = (x - \bar{\xi}_1)Q(x) + b_n.$$

Koeficienty $b_0,\ldots,b_{n-1}$ polynomu Q lze získat Hornerovým schématem. V případě, že ξ_1 je přesný kořen polynomu P, je $b_n = P(\xi_1) = 0$. Je-li $\bar{\xi}_1$ aproximací kořene ξ_1, je $P(\bar{\xi}_1) = b_n \neq 0$.

Hornerova schematu lze použít i pro výpočet hodnoty derivace polynomu P v bodě α. Derivací vztahu $P(x) = (x - \alpha)Q(x) + b_n$ dostaneme

$$P'(x) = Q(x) + (x - \alpha)Q'(x)$$

a tedy $P'(\alpha) = Q(\alpha)$.

$$\begin{aligned} Q(x) &= (x - \alpha)R(x) + c_{n-1} \\ R(x) &= c_0 x^{n-2} + \ldots c_{n-2} \\ \text{a } Q(\alpha) &= c_{n-1}. \end{aligned}$$

Výpočet lze vhodně uspořádat do tabulky

	a_0	a_1	a_2	$\ldots$	a_{n-1}	a_n
α		αb_0	αb_1	$\ldots$	αb_{n-2}	αb_{n-1}
	b_0	b_1	b_2	$\ldots$	b_{n-1}	$b_n = P(\alpha)$
α		αc_0	αc_1	$\ldots$	αc_{n-2}	
	c_0	c_1	c_2	$\ldots$	$c_{n-1} = Q(\alpha) = P'(\alpha)$	

Obdobným způsobem lze vypočítat hodnoty $P^{(j)}(\alpha)$ pro libovolné $j \geq 1$. Položme

$$Q^{(0)}(x) = Q(x) = b_0^{(0)} x^{n-1} + \ldots + b_{n-1}^{(0)}$$

takže

$$P(x) = (x - \alpha)Q^{(0)}(x) + b_n^{(0)}$$

Dále

$$Q^{(0)}(x) = (x - \alpha)Q^{(1)}(x) + b_{n-1}^{(1)}$$

pro

$$Q^{(1)}(x) = b_0^{(1)} x^{n-1} + \ldots + b_{n-2}^{(1)},$$

a tedy

$$P(x) = Q^{(1)}(x)(x - \alpha)^2 + b_{n-1}^{(1)}(x - \alpha) + b_n^{(0)}.$$

Dalším postupem bychom dostali vyjádření

$$P(x) = b_0^{(n)}(x - \alpha)^n + b_1^{(n-1)}(x - \alpha)^{n-1} + \ldots + b_{n-1}^{(1)}(x - \alpha) + b_n^{(0)}, \qquad (3.6)$$

kde koeficienty $b_i^{(n-i)}$, $i = 0,\ldots,n$ dostaneme postupným použitím Hornerova schematu podle následující tabulky

	a_0	a_1	a_2	$\dots$	a_{n-2}	a_{n-1}	a_n
α		$\alpha b_0^{(0)}$	$\alpha b_1^{(0)}$	$\dots$	$\alpha b_{n-3}^{(0)}$	$\alpha b_{n-2}^{(0)}$	$\alpha b_{n-1}^{(0)}$
	$b_0^{(0)}$	$b_1^{(0)}$	$b_2^{(0)}$	$\dots$	$b_{n-2}^{(0)}$	$b_{n-1}^{(0)}$	$b_n^{(0)}$
α		$\alpha b_0^{(1)}$	$\alpha b_1^{(1)}$	$\dots$	$\alpha b_{n-3}^{(1)}$	$\alpha b_{n-2}^{(1)}$	
	$b_0^{(1)}$	$b_1^{(1)}$	$b_2^{(1)}$	$\dots$	$b_{n-2}^{(1)}$	$b_{n-1}^{(1)}$	
α		$\alpha b_0^{(2)}$	$\alpha b_1^{(2)}$	$\dots$	$\alpha b_{n-3}^{(2)}$		
	$b_0^{(2)}$	$b_1^{(2)}$	$b_2^{(2)}$	$\dots$	$b_{n-2}^{(2)}$		
	$\vdots$						
	$b_0^{(n-2)}$	$b_1^{(n-2)}$	$b_2^{(n-2)}$				
α		$\alpha b_0^{(n-1)}$					
	$b_0^{(n-1)}$	$b_1^{(n-1)}$					
α							
	$b_0^{(n)}$						

Přitom je z (3.6) zřejmé, že $P^{(j)}(\alpha) = j!\, b_{n-j}^{(j)}$.

Jak víme z předchozí kapitoly, pro konvergenci Newtonovy metody je třeba znát dostatečně dobrou počáteční aproximaci. Avšak neexistuje žádné pravidlo, které by zaručovalo dobrou počáteční aproximaci pro libovolný polynom. Na druhé straně takové pravidlo platí pro speciální případ, kdy všechny kořeny $\xi_i, i = 1, \dots, n$ jsou reálné a platí

$$\xi_1 \geq \xi_2 \geq \dots \geq \xi_n.$$

Toto pravidlo je v podstatě obdobou věty 2.10 a zní takto:

Věta 3.5. *Nechť $P \in \Pi_n$ je polynom stupně $n \geq 2$. Nechť všechny kořeny ξ_i,*

$$\xi_1 \geq \xi_2 \geq \dots \geq \xi_n,$$

jsou reálné. Pak posloupnost $\{x^k\}_{k=0}^{\infty}$ určená Newtonovou metodou je konvergentní klesající posloupnost pro každou počáteční aproximaci $x^0 > \xi_1$, $\xi_1 = \lim_{k\to\infty} x^k$.

Důkaz. Bez újmy na obecnosti lze předpokládat, že $P(x^0) > 0$. Protože P nemění znaménko pro $x > \xi_1$, máme

$$P(x) > 0 \quad \text{pro} \quad x > \xi_1$$

a tedy $a_0 > 0$. Derivace P' má $n-1$ reálných kořenů α_i s vlastností (v důsledku Rolleovy věty):

$$\xi_1 \geq \alpha_1 \geq \xi_2 \geq \alpha_2 \geq \dots \geq \alpha_{n-1} \geq \xi_n.$$

Protože P' je polynom stupně nejvýše $n-1$, jsou toto všechny jeho kořeny a $P'(x) > 0$ pro $x > \alpha_1$, neboť $a_0 > 0$. Opětovnou aplikací Rolleovy věty dostaneme

$$P''(x) > 0 \text{ pro } x > \alpha_1 \qquad (n \geq 2),$$
$$P'''(x) \geq 0 \text{ pro } x \geq \alpha_1.$$

Tedy P a P' jsou konvexní funkce pro $x \geq \alpha_1$. Nyní na intervalu $[\alpha_1, x^0]$ lze aplikovat větu (2.9). Znaménko polynomu P v bodě x^0 je stejné jako znaménko P'' na intervalu $[\alpha_1, x^0]$ a tedy pro počáteční aproximaci $x^0 > \xi_1$ posloupnost $\{x^k\}_{k=0}^{\infty}$ určená Newtonovou metodou konverguje monotonně ke kořenu ξ_1. □

Poznámka 3. Pro volbu počáteční aproximace lze užít odhadů uvedených ve větách 3.1, 3.2.

Newtonova metoda konverguje kvadraticky, ale tato konvergence nemusí vždy znamenat rychlou konvergenci. Jestliže počáteční aproximace x^0 leží daleko od kořene, pak posloupnost $\{x^k\}_{k=0}^{\infty}$ určená Newtonovou metodou může konvergovat pomalu, neboť pro x^k velké

$$x^{k+1} = x^k - \frac{(x^k)^n + \dots}{n(x^k)^{n-1} + \dots} \approx x^k \left(1 - \frac{1}{n}\right),$$

odkud je vidět, že je malý rozdíl mezi x^k, x^{k+1}. Z tohoto důvodu se budeme zabývat následující *zdvojenou metodou*.

$$x^{k+1} = x^k - 2\frac{P(x^k)}{P'(x^k)}, \qquad k = 0, 1, 2, \dots \tag{3.7}$$

místo přímé Newtonovy metody.

Věta 3.6. *Nechť $P \in \Pi_n$, $n \geq 2$, a nechť všechny kořeny $\xi_i, i = 1, \dots, n$ polynomu P jsou reálné a $\xi_1 \geq \xi_2 \geq \dots \geq \xi_n$. Nechť α_1 je největší kořen P':*

$$\xi_1 \geq \alpha_1 \geq \xi_2.$$

Pro $n = 2$ předpokládejme $\xi_1 > \xi_2$. Pak pro každé $z > \xi_1$ jsou čísla

$$z' = z - \frac{P(z)}{P'(z)}, \qquad y = z - 2\frac{P(z)}{P'(z)}, \qquad y' = y - \frac{P(y)}{P'(y)} \tag{3.8}$$

definována a platí

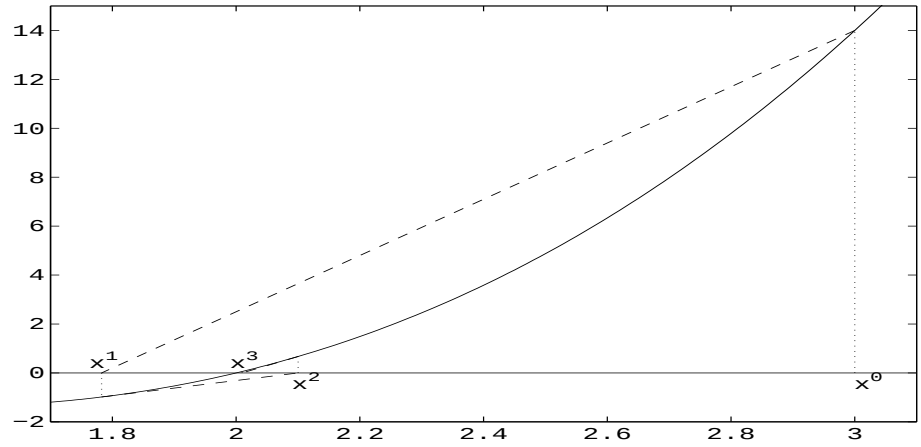
$$\begin{aligned} \alpha_1 &< y, \\ \xi_1 \leq y' &\leq z'. \end{aligned} \tag{3.9}$$

Důkaz lze najít v [5].

Obrázek 3.1 ilustruje geometrický význam této metody.

Metoda (3.7) není z geometrického hlediska založena na průsečíku tečny s osou x, ale na průsečíku sečny, která má směrnici $P'(x)/2$, s osou x. To znamená, že při použití této metody můžeme „přestřelit“ bod ξ_1, tj. pro bod y ve vztahu (3.8) může platit $y < \xi_1$. Ale z věty 3.6 plyne, že tento bod splňuje rovněž nerovnost (3.9) a použitím klasické Newtonovy metody dostaneme bod y', který je lepší aproximací ξ_1 než bod z'.

Praktický význam věty 3.6 je následující:



Obr. 3.1: Zdvojená Newtonova metoda, $P(x) = x^3 + x^2 - 10x + 8$, $x^0 = 3$

Začneme s počáteční aproximací $x^0 > \xi_1$. Pro posloupnost $\{x^k\}_{k=0}^{\infty}$ generovanou zdvojenou metodou

$$x^{k+1} = x^k - 2\frac{P(x^k)}{P'(x^k)}, \qquad k = 0, 1, \ldots$$

mohou nastat dva případy:

(a) $P(x^0)P(x^k) > 0$ pro všechna k. V tomto případě

$$x^0 > x^1 > \ldots > x^k > \ldots \geq \xi_1, \qquad \lim_{k\to\infty} x^k = \xi_1$$

a konvergence je rychlejší než konvergence přímé Newtonovy metody.

(b) Existuje x^{k_0} tak, že $P(x^{k_0})P(x^0) < 0$, $P(x^k)P(x^0) > 0$ pro $0 \leq k < k_0$. V tomto případě tedy došlo k „přestřelení“ bodu ξ_1 a platí

$$x^0 > x^1 > \ldots > x^{k_0-1} > \xi_1 > y = x^{k_0} > \alpha_1 > \xi_2.$$

Položme $y^0 = x^{k_0}$ a pokračujme dále klasickou Newtonovou metodou s touto počáteční aproximací:

$$y^{k+1} = y^k - \frac{P(y^k)}{P'(y^k)}, \qquad k = 0, 1, 2, \ldots$$

Dostaneme opět monotonní posloupnost

$$y^1 > y^2 > \ldots > y^k > \ldots \geq \xi_1, \qquad \lim_{k\to\infty} y^k = \xi_1.$$

Příklad 3.4. Ukážeme si rozdíl mezi rychlostí konvergence přímé a zdvojené Newtonovy metody. Položme $P(x) = \prod_{i=1}^{8}(x-i)$, tedy $\xi_1 = 8$, $\xi_2 = 7,\ldots,\xi_8 = 1$. Počáteční aproximaci zvolme $x_0 = 20$. Jednotlivé iterace dává následující tabulka:

	Newtonova metoda	zdvojená Newtonova metoda
x^0	20,0	20,0
x^1	18,105567	16,211133
x^2	16,454192	13,398883
x^3	15,016438	11,329903
x^4	13,766710	9,834383
x^5	12,682811	8,794966
x^6	11,745573	8,148323
x^7	10,938548	7,929357
x^8	10,247782	8,016696
x^9	9,661673	8,000686
x^10	9,170955	8,000001

U zdvojené Newtonovy metody došlo k „přestřelení“ kořene ξ_1 u aproximace x^7, pro následující iterace už je použita přímá Newtonova metoda.

Výše uvedeným postupem nalezneme aproximaci $\tilde{\xi}_1$ největšího kořene ξ_1 polynomu P. Pro určení dalších kořenů se nabízí jednoduchá myšlenka: známe aproximaci $\tilde{\xi}_1$, vydělíme polynom P dvojčlenem $(x-\tilde{\xi}_1)$ a výsledný polynom

$$P_1(x) = \frac{P(x)}{x-\tilde{\xi}_1}$$

je polynom stupně $n-1$ s největším kořenem ξ_2. Tato metoda se nazývá metoda *snižování stupně*. Takto bychom mohli teoreticky najít všechny kořeny polynomu P. Ale praktická realizace tohoto procesu není bez problémů. Kořen ξ_1 je znám přibližně a vzhledem k zaokrouhlovacím chybám nelze přesně určit polynom P_1 (viz poznámka 2). Polynom, který získáme uvedeným dělením bude mít tedy kořeny odlišné od $\xi_2,\ldots,\xi_n$. Při dalším opakování tohoto postupu mohou být poslední kořeny polynomu P určeny zcela nepřesně. Z těchto důvodů byly navrženy různé modifikace metody snižování stupně ([5]).

Metodu, která se vyhýbá přímému snižování stupně, navrhl v roce 1954 Maehly ([5]). Základní myšlenka této metody spočívá ve vhodném vyjádření derivace polynomu nižšího stupně:

$$P_1'(x) = \frac{P'(x)}{x-\tilde{\xi}_1} - \frac{P(x)}{(x-\tilde{\xi}_1)^2}, \tag{3.10}$$

Dosazením tohoto vyjádření do vzorce pro Newtonovu metodu pro polynom P_1 dostaneme:

$$x^{k+1} = x^k - \frac{P_1(x^k)}{P_1'(x^k)} = x^k - \frac{P(x^k)}{P'(x^k) - \dfrac{P(x^k)}{x^k - \tilde{\xi}_1}} \tag{3.11}$$

Obecně, jestliže jsme již nalezli aproximace kořenů $\tilde{\xi}_1, \ldots, \tilde{\xi}_j$, postupujeme obdobně a sestrojíme polynom

$$P_j(x) = \frac{P(x)}{(x - \tilde{\xi}_1) \ldots (x - \tilde{\xi}_j)},$$

$$P_j'(x) = \frac{P'(x)}{(x - \tilde{\xi}_1) \ldots (x - \tilde{\xi}_j)} - \frac{P(x)}{(x - \tilde{\xi}_1) \ldots (x - \tilde{\xi}_j)} \sum_{i=1}^{j} \frac{1}{x - \tilde{\xi}_i}$$

Newtonova metoda pro nalezení kořene ξ_{j+1} je tvaru

$$x^{k+1} = \Phi_j(x^k), \qquad \Phi_j(x) = x - \frac{P(x)}{P'(x) - \displaystyle\sum_{i=1}^{j} \frac{P(x)}{x - \tilde{\xi}_i}}. \tag{3.12}$$

Přednost této metody spočívá ve skutečnosti, že posloupnost $\{x^k\}$ generovaná metodou (3.12) konverguje kvadraticky ke kořenu ξ_{j+1} i v případě, že $\tilde{\xi}_1, \ldots, \tilde{\xi}_j$ nejsou kořeny P (konvergence je pouze lokální v tomto případě). Tedy výpočet ξ_{j+1} není citlivý na chyby při výpočtu předchozích kořenů. Můžeme také vhodně aplikovat zdvojenou Newtonovu metodu.

Příklad 3.5. Užitím zdvojené Newtonovy–Maehlyovy metody nalezněte kořeny polynomu $P(x) = x^3 + x^2 - 10x + 8$.
Řešení. Nechť $x^0 = 3$, tj. počáteční aproximace kořene ξ_1 je $\xi_1^0 = 3$.

Aproximace kořene ξ_1: $\xi_1^1 = 1{,}782608695652$
$\xi_1^2 = 2{,}10014059474224$
$\xi_1^3 = 2{,}00971540717739$
$\xi_1^4 = 2{,}0001079735567$
$\xi_1^5 = 2{,}00000001359833$

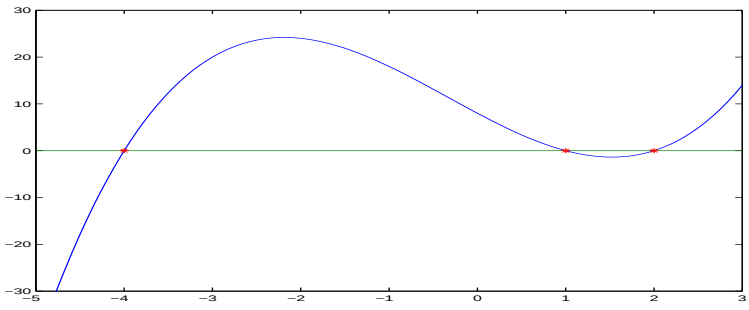
Hned první aproximace ξ_1^1 „přestřelila“ hledaný kořen $\xi_1 = 2$, tj. $\xi_1^0 > \xi_1$, $\xi_1^1 < \xi_1$, pro výpočet dalších aproximací je tedy použita klasická Newtonova metoda.
Dále, položme $\xi_2^0 = 1{,}9$ (což je počáteční aproximace pro kořen ξ_2).

Je $\xi_2^1 = 0{,}33823529411765$
$\xi_2^2 = 1{,}11911764705882$
$\xi_2^3 = 1{,}00270873930706$
$\xi_2^4 = 1{,}00000146586547$
$\xi_2^5 = 1{,}00000000000043$

Opět první aproximace ξ_1^1 je menší než hledaný kořen $\xi_2 = 1$, takže pro výpočet dalších aproximací použijeme klasickou Newtonovu metodu.
Za počáteční aproximace kořene ξ_3 zvolíme $\xi_3^0 = 0{,}9$. Další aproximace jsou

$$\begin{aligned}\xi_3^1 &= -8{,}899999999935\\ \xi_3^2 &= -4{,}000000000\ldots\\ \xi_3^2 &= \xi_3^3 = \xi_3^4\end{aligned}$$

Pro srovnání: přesné kořeny jsou $\xi_1 = 2$, $\xi_2 = 1$, $\xi_3 = -4$. Tvar polynomu zachycuje obrázek 5. Pro srovnání uveďme výsledky, které bychom dostali metodou



Obr. 3.2: Průběh polynomu $P(x) = x^3 + x^2 - 10x + 8$.

snižování stupně. Výpočet ξ_1 by samozřejmě probíhal stejným způsobem. Pokud $\xi_1^5 = 2{,}00000001359833$ považujeme za dostatečně přesnou aproximaci ξ_1, dostaneme polynom $P_1(x) = P(x)/(x - \xi_1^5)$ ve tvaru

$$P_1(x) = x^2 + 3.00000001359833\,x - 3.99999993200835$$

a následným použitím zdvojené Newtonovy metody vyjde $\xi_2^5 = 0.99999998368243$. Dalším snížením stupně polynomu dostaneme $P_2(x) = x + 3.99999999728076$, takže nelze dostat přesnější aproximaci kořene ξ_3 než hodnotu -3.99999999728076.

Význam Maehlyovy metody se projevuje zejména u polynomů vyšších stupňů, jak dokládá následující příklad.

Příklad 3.6. Položme $P(x) = \prod_{i=1}^{15}(x-i)$. V uvedené tabulce vidíme rozdíl mezi metodou snižování stupně a Maehlyovou metodou. V tabulce nejsou uvedeny všechny výsledky, jen aproximace nejmenších kořenů, kde je chyba největší.

	snižování řádu	Maehlyova metoda
ξ_1	15,0000037	15,0000037
	$\vdots$	
ξ_{13}	2,92420696801	2,999999999978
ξ_{14}	2,01522846251	2,000000000007
ξ_{15}	0,99878713149	1,000000000000

§ 3.4. Bairstowova metoda

Doposud jsme se zabývali výpočtem reálných kořenů polynomu. Pokud jde o výpočet komplexních kořenů, lze užít některých z předcházejících metod, např. Newtonovy metody s komplexní počáteční aproximací, nebo některých speciálních metod: Lehmerovy-Schurovy metody, Graeffovy metody, Bernoulliovy metody nebo Bairstowovy metody. Pojednáme zde podrobněji o Bairstowově metodě. Ostatní zmíněné metody jsou uvedeny např. v [19].

Podstatou Bairstowovy metody je myšlenka nalezení kvadratického trojčlenu, který je dělitelem daného polynomu:

Nechť

$$P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n.$$

Označme z, $\bar{z}$, $z = u + \mathrm{i}v$, dvojici komplexně sdružených kořenů polynomu P. Čísla z, $\bar{z}$ lze považovat za kořeny kvadratického trojčlenu $D(x) = x^2 + px + q$, $p = -2u$, $q = u^2 + v^2$. Naším úkolem je najít čísla p, q tak, aby polynom D dělil polynom P beze zbytku. Budeme-li znát čísla p, q, pak snadno určíme komplexní kořeny $z, \bar{z}$ polynomu P. Tuto myšlenku lze formálně zapsat takto:

$$P(x) = D(x)Q(x) + Ax + B, \tag{3.13}$$

kde

$$\begin{aligned} D(x) &= x^2 + px + q, \\ Q(x) &= Q(x, p, q) \qquad \text{je polynom stupně } n-2, \\ A &= A(p, q), \\ B &= B(p, q). \end{aligned}$$

Je třeba určit p, q tak, aby

$$A(p, q) = 0, \qquad B(p, q) = 0. \tag{3.14}$$

Systém (3.14) je systém nelineárních rovnic a budeme ho řešit Newtonovou metodou pro systémy nelineárních rovnic (2.47).

Považujeme-li kvadratický trojčlen $D(x) = x^2 + px + q$ za aproximaci dělitele, dostaneme další aproximaci $D_1(x) = x^2 + p_1 x + q_1$, $p_1 = p + h$, $q_1 = q + k$, řešením soustavy

$$\begin{pmatrix} \dfrac{\partial A}{\partial p} & \dfrac{\partial A}{\partial q} \\ \dfrac{\partial B}{\partial p} & \dfrac{\partial B}{\partial q} \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = - \begin{pmatrix} A(p,q) \\ B(p,q) \end{pmatrix}$$

neboli, označíme-li $\dfrac{\partial A}{\partial p} = A'_p$, $\dfrac{\partial A}{\partial q} = A'_q$, $\dfrac{\partial B}{\partial p} = B'_p$, $\dfrac{\partial B}{\partial q} = B'_q$,

$$\begin{aligned} A(p,q) + A'_p(p,q)h + A'_q(p,q)k &= 0, \\ B(p,q) + B'_p(p,q)h + B'_q(p,q)k &= 0. \end{aligned} \tag{3.15}$$

Veličiny $A(p,q)$, $B(p,q)$ lze získat zobecněným Hornerovým schematem (viz strana 88) při dělení polynomu P trojčlenem D. Derivujeme vztah (3.13) podle p a q:

$$\begin{aligned} 0 &= xQ(x) + Q'_p(x)D(x) + A'_p x + B'_p \\ 0 &= \ Q(x) + Q'_q(x)D(x) + A'_q x + B'_q \end{aligned}$$

Odtud

$$\begin{aligned} &\text{(a)} \qquad xQ(x) = -Q'_p(x)D(x) - A'_p x - B'_p, \\ &\text{(b)} \qquad Q(x) = -Q'_q(x)D(x) - A'_q x - B'_q. \end{aligned} \tag{3.16}$$

Je zřejmé, že $-A'_p$, $-B'_p$ resp. $-A'_q$, $-B'_q$ jsou koeficienty lineárních zbytků při dělení polynomu $xQ(x)$ polynomem $D(x)$, resp. $Q(x)$ polynomem $D(x)$. Položme

$$a = -A'_q, \qquad b = -B'_q.$$

Tato čísla lze opět získat zobecněným Hornerovým algoritmem pro dělení polynomů $Q(x)/D(x)$. Vypočteme nyní A'_p, B'_p.

Vynásobme (3.16b) číslem x:

$$xQ(x) = -xQ'_q(x)D(x) + ax^2 + bx$$

a po úpravě můžeme tento vztah zapsat ve tvaru

$$xQ(x) = a(x^2 + px + q) + bx - xQ'_q(x)D(x) - apx - q,$$

a tedy

$$xQ(x) = \big(a - xQ'_q(x)\big)\, D(x) + (b - ap)x - aq. \tag{3.17}$$

Porovnáním (3.17) a (3.16a) dostaneme

$$A'_p = ap - b, \qquad B'_p = aq.$$

Soustavu (3.15) můžeme nyní zapsat takto:

$$\begin{aligned}(ap-b)h - ak + A &= 0,\\ aqh - bk + B &= 0.\end{aligned} \tag{3.18}$$

Vyřešením této soustavy získáme čísla h, k a kvadratický trojčlen $D_1(x) = x^2 + (p+h)x + q + k$, jehož kořeny jsou aproximací kořenů z, $\bar{z}$ polynomu P. Postup opakujeme. Jako kriterium pro zastavení výpočtu lze zvolit: $|h| < \varepsilon|p|$, $|k| < \varepsilon|q|$, $\varepsilon > 0$ zadaná přesnost. Na závěr této kapitoly uvedeme zobecněné Hornerovo schéma pro výpočet hodnot A, B, a, b.

Zobecněné Hornerovo schema.
Je dán polynom $P(x) = a_0x^n + \ldots + a_n$ a aproximace kvadratického trojčlenu $D(x) = x^2 + px + q$. Hodnoty A, B, a, b vypočítáme zobecněným Hornerovým schematem. Výpočet lze uspořádat do tabulky:

	a_0	a_1	a_2	$\ldots$	a_{n-3}	a_{n-2}	a_{n-1}	a_n
$-p$		$-pb_0$	$-pb_1$	$\ldots$	$-pb_{n-4}$	$-pb_{n-3}$	$-pb_{n-2}$	
$-q$			$-qb_0$	$\ldots$	$-qb_{n-5}$	$-qb_{n-4}$	$-qb_{n-3}$	$-qb_{n-2}$
	b_0	b_1	b_2	$\ldots$	b_{n-3}	b_{n-2}	A	B
$-p$		$-pc_0$	$-pc_1$	$\ldots$	$-pc_{n-4}$			
$-q$			$-qc_0$	$\ldots$	$-qc_{n-5}$	$-qc_{n-4}$		
	c_0	c_1	c_2		a	b		

Zde $Q(x) = b_0x^{n-2} + \ldots + b_{n-2}$,

$$\begin{aligned}b_j &= a_j - pb_{j-1} - qb_{j-2}, \qquad j = 2, \ldots, n-2\\ b_0 &= a_0\\ b_1 &= a_1 - pb_0\\ A &= a_{n-1} - pb_{n-2} - qb_{n-3}\\ B &= a_n - qb_{n-2}\end{aligned}$$

Dále

$$Q(x) = D(x)R(x) + ax + b,$$

kde $R(x) = c_0x^{n-4} + \ldots + c_{n-4}$. Koeficienty c_j, a, b jsou určeny obdobným způsobem jako koeficienty b_j, A, B, a to:

$$a = -pc_{n-4} - qc_{n-5} + b_{n-3}, \qquad b = -qc_{n-4} + b_{n-2},$$

$$c_j = b_j - pc_{j-1} - qc_{j-2},\ j = 2, \ldots, n-4\,.$$

Příklad 3.7. Užitím Bairstowovy metody nalezněte dvojici komplexně sdružených kořenů polynomu

$$P(x) = x^4 - 3x^2 + 4x - 1.$$

Řešení. Ze Sturmovy věty plyne, že polynom má 2 reálné a 2 komplexně sdružené kořeny.

Za počáteční aproximaci kvadratického trojčlenu zvolme $D_0(x) = x^2 + x + 1$, tj. $p = 1$, $q = 1$.

Aplikujme nyní zobecněné Hornerovo schema:

	1	0	−3	4	−1
−1		−1	1	3	
−1			−1	1	3
	1	−1	−3	8	2
−1		−1			
−1			−1		
	1	−2	−4		

Odtud plyne, že $A = 8$, $B = 2$, $a = -2$, $b = -4$. Odpovídající systém rovnic (3.18) je tvaru

$$\begin{aligned} 2h + 2k + 8 &= 0 \\ -2h + 4k + 2 &= 0. \end{aligned}$$

Odtud $h = -7/3$, $k = -5/3$ a tedy nový kvadratický trojčlen je tvaru

$$D_1(x) = x^2 + \left(1 - \tfrac{7}{3}\right) x + \left(1 - \tfrac{5}{3}\right) = x^2 - \tfrac{4}{3}x - \tfrac{2}{3}.$$

Nyní výpočet opakujeme s tímto kvadratickým trojčlenem. Koeficienty kvadratických trojčlenů jsou uvedeny v následující tabulce.

	p_k	q_k
$k = 0$	1	1
$k = 1$	−4/3	−2/3
$k = 2$	−2,283000949	2,1946816134
$k = 3$	−2,03645296288	1,53678222972
⋮		
$k = 8$	−1,90640113643838	1,3662797903433
$k = 9$	−1,90640113643838	1,3662797903433

Aproximace 8, 9 ukazují, že výpočet je stabilizován, neboť absolutní hodnota rozdílu dvou po sobě jdoucích aproximací je zanedbatelná. Za aproximaci dvojčlenu lze vzít

$$D(x) \approx x^2 - 1{,}90640113643838x + 1{,}3662797903433.$$

Kořeny tohoto dvojčlenu jsou

$$\xi_{1,2} = 0{,}95320056821919 \pm 0{,}67652677240516\mathrm{i}.$$

Cvičení ke kapitole 3

1. Pomocí Hornerova schematu vypočtěte hodnotu $P(\alpha)$, $P'(\alpha)$ a $P''(\alpha)$ pro polynom $P(x) = 2x^5 - x^4 + 3x^2 + x - 5$ a $\alpha = 2$.

2. Užitím zobecněného Hornerova algoritmu vypočtěte hodnotu polynomu z příkladu 1 v bodě $z = 1 + \mathrm{i}$.
(Řešení: $P(1 + \mathrm{i}) = -8 - \mathrm{i}$.)

3. Je dána rovnice $3x^3 - x - 1 = 0$. Určete interval, ve kterém leží kladný kořen této rovnice.

4. Užitím Sturmovy věty určete počet reálných kořenů polynomů

   a) $P(x) = x^4 - 4x + 1$ (2 reálné kořeny: $\xi_1 \in [0, 1]$, $\xi_2 \in [1, 2]$),

   b) $P(x) = x^3 + 3x^2 - 1$ (3 reálné kořeny: $\xi_1 \in [0, 1]$, $\xi_2 \in [-1, 0]$, $\xi_3 \in [-3, -2]$).

5. Sestrojte Sturmovu posloupnost pro polynom

   $$P(x) = x^3 + 3x^2 - 1$$

   a určete počet reálných kořenů tohoto polynomu. Ukažte, že všechny reálné kořeny leží v intervalu $[-3, 1]$ a najděte intervaly, ve kterých leží vždy právě jeden kořen.

6. Pomocí Sturmovy věty určete počet reálných kořenů polynomu $P(x) = x^4 - x^2 + 3$.

7. Užitím zdvojené Newtonovy-Maehlyovy metody nalezněte kořeny polynomu $P(x) = x^3 + 3x^2 - 1$.
(Řešení: $\xi_1 = 0{,}532089$, $\xi_2 = -0{,}652706$, $\xi_3 = -2{,}87938$.)

8. Určete reálné a komplexní kořeny polynomu

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

   (Řešení: $\xi_1 = 1{,}2412$, $\xi_2 = -0{,}876053$, $\xi_{3,4} = -0{,}124035 \pm 1{,}74096\mathrm{i}$.)

9. Užitím Bairstowovy metody nalezněte komplexní kořeny polynomu

   $$P(x) = x^4 + 4x^2 - 3x - 1.$$

   (Řešení: $\xi_{1,2} = -0{,}3111 \pm 2{,}1231\mathrm{i}$.)

Kontrolní otázky ke kapitole 3

1. Platí věta 3.1 i pro polynomy s komplexními koeficienty?
2. Bylo by možné použít zdvojenou Newtonovu metodu v případě $n = 2$ a $\xi_1 = \xi_2$ (viz předpoklady věty 3.6)?
3. Co by se mohlo stát, pokud by při použití zdvojené Newtonovy metody nebyl splněn předpoklad, že všechny kořeny jsou reálné?
4. Lze použít Maehlyovu metodu pro polynom s vícenásobnými kořeny?
5. Jak se dá použít princip Maehlyovy metody v případě, že polynom má komplexní kořeny?

Kapitola 4

Přímé metody řešení systémů lineárních rovnic

Nyní se budeme zabývat metodami pro řešení systému lineárních rovnic

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{4.1}$$

Předpokládáme, že A je reálná čtvercová matice řádu n ($A \in \mathcal{M}_n$), $\boldsymbol{b}$ je reálný vektor. Je-li matice A regulární, má systém (4.1) jediné řešení $\boldsymbol{x}^*$:

$$\boldsymbol{x}^* = A^{-1}\boldsymbol{b} \tag{4.2}$$

V této kapitole se budeme zabývat přímými metodami řešení systému (4.1), tj. metodami, jejichž aplikací získáme přesné řešení $\boldsymbol{x}^*$ po konečném počtu kroků za předpokladu, že všechny aritmetické operace provádíme přesně a vstupní údaje jsou dány přesně. Při vyšetřování těchto metod se budeme také zabývat otázkou, kolik aritmetických operací je třeba pro realizaci výpočtu užitím dané metody. Jelikož výpočty provádíme v důsledku zápisů čísel v počítači pouze s přibližnými čísly a zaokrouhlujeme během výpočtu, budeme věnovat také pozornost přesnosti nalezeného řešení užitím dané metody — to jsou tzv. a priori odhady.

§ 4.1. Systémy lineárních rovnic

Připomeňme nyní základní poznatky z lineární algebry.

Matici tvaru

$$(A \mid \boldsymbol{b}) = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ a_{21} & \cdots & a_{2n} & b_2 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & b_n \end{array}\right) \tag{4.3}$$

nazýváme *rozšířenou* maticí systému (4.1).

Definice 4.1. Systém (4.1) se nazývá *řešitelný* (resp. *neřešitelný*), jestliže existuje alespoň jedno (resp. neexistuje žádné) řešení.

Věta 4.1. (Frobenius). *Systém lineárních rovnic* (4.1) *je řešitelný právě tehdy, když hodnost matice A je rovna hodnosti rozšířené matice systému* $(A \mid \boldsymbol{b})$.

Definice 4.2. Matice $R \in \mathcal{M}_n$, $R = (r_{ij})$, se nazývá *horní trojúhelníková* matice, jestliže $r_{ij} = 0$ pro $i > j$.

Matice $R \in \mathcal{M}_n$, $R = (r_{ij})$, se nazývá *dolní trojúhelníková* matice, jestliže $r_{ij} = 0$ pro $i < j$.

Definice 4.3. Matice $A \in \mathcal{M}_n$ se nazývá *pásová*, jestliže existují přirozená čísla p, q, $1 < p$, $q < n$ taková, že $a_{ij} = 0$, jestliže $i + p \leq j$ nebo $j + q \leq i$. Šířka pásu $w = p + q - 1$.

Poznámka 1. Pro $p = q = 2$ se pásová matice nazývá *třídiagonální* a je tvaru

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & \cdots & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & & & \vdots \\ 0 & a_{32} & a_{33} & & & \vdots \\ \vdots & & & \ddots & & 0 \\ \vdots & & & & a_{n-1,n-1} & a_{n-1,n} \\ 0 & \cdots & \cdots & 0 & a_{n,n-1} & a_{nn} \end{pmatrix}.$$

Definice 4.4. Matice $A \in \mathcal{M}_n$ se nazývá *ryze řádkově diagonálně dominantní*, jestliže

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \qquad i = 1, \ldots, n. \tag{4.4}$$

Věta 4.2. *Jestliže $A \in \mathcal{M}_n$ je ryze řádkově diagonálně dominantní, je regulární.*

Důkaz. Předpokládejme, že A není regulární. Pak systém $A\boldsymbol{x} = \boldsymbol{o}$ má netriviální řešení $\tilde{\boldsymbol{x}} = (x_1, \ldots, x_n)^T$. Nechť $|x_k| = \max\limits_{1 \leq i \leq n} |x_i|$. Podle předpokladu je $x_k \neq 0$.

Nyní, i-tá rovnice systému $A\boldsymbol{x} = \boldsymbol{o}$ je tvaru

$$\sum_{j=1}^{n} a_{ij} x_j = 0$$

a k-tá rovnice může být zapsána ve tvaru

$$a_{kk} x_k + \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} x_j = 0,$$

tj.

$$a_{kk} = - \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} \frac{x_j}{x_k}.$$

Přechodem k absolutním hodnotám dostaneme

$$|a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}| \frac{|x_j|}{|x_k|} \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}|,$$

což je spor s předpokladem (4.4). Matice A musí tedy být regulární. □

Definice 4.5. Symetrická matice $A \in \mathcal{M}_n$ se nazývá *pozitivně definitní*, jestliže $\boldsymbol{x}^T A \boldsymbol{x} > 0$ pro každý nenulový vektor $\boldsymbol{x} \in \mathbb{R}^n$.

Věta 4.3. *Pozitivně definitní matice je regulární.*

Důkaz. Předpokládejme, že A je singulární, tzn. že systém $A\boldsymbol{x} = \boldsymbol{o}$ má netriviální řešení $\boldsymbol{x}' \neq \boldsymbol{o}$, $A\boldsymbol{x}' = \boldsymbol{o}$. Pro tento vektor $\boldsymbol{x}'$ platí $\boldsymbol{x}'^T A \boldsymbol{x}' = 0$, což je spor s pozitivní definitností matice A a tedy A je regulární. □

Většina výpočetních algoritmů numerické lineární algebry má společnou základní strukturu, kterou lze popsat takto:

1. Daný problém se převede na „redukovaný“ problém.
2. Řeší se tento redukovaný problém při využití jeho speciální struktury.
3. Řešení původního problému se zpětně získá z řešení redukovaného problému.

Úmluva. V celé této kapitole budeme předpokládat, že $A \in \mathcal{M}_n$. Tuto skutečnost nebudeme zdůrazňovat, pokud nemůže dojít k nedorozumění.

§ 4.2. Gaussova eliminační metoda

Nejznámější přímou metodou pro řešení systému lineárních rovnic je Gaussova eliminační metoda „GEM“, kterou lze rovněž zahrnout do výše uvedené třídy algoritmů. Hlavní myšlenka této metody spočívá v převedení daného systému $A\boldsymbol{x} = \boldsymbol{b}$ vhodnými ekvivalentními úpravami na systém $R\boldsymbol{x} = \boldsymbol{c}$ s horní trojúhelníkovou maticí R, tj. problém se převede na redukovaný problém. Této etapě říkáme *přímý chod*. Tento systém má stejné řešení jako původní systém $A\boldsymbol{x} = \boldsymbol{b}$ a jeho řešení lze snadno získat zpětnou substitucí (za předpokladu $r_{ii} \neq 0$, $i = 1, \ldots, n$):

$$x_i = \frac{c_i - \sum_{k=i+1}^{n} r_{ik} x_k}{r_{ii}}, \qquad i = n, n-1, \ldots, 1. \tag{4.5}$$

Jelikož systém $R\boldsymbol{x} = \boldsymbol{c}$ řešíme od poslední rovnice, říká se této etapě *zpětný chod*.

V prvém kroku algoritmu se vhodný násobek první rovnice odečítá od zbývajících $n-1$ rovnic tak, aby koeficienty u x_1 ve zbývajících rovnicích byly rovny nule; tedy x_1 zůstává pouze v první rovnici. Tento postup je možný pouze za předpokladu, že $a_{11} \neq 0$. Splnění tohoto předpokladu lze dosáhnout vhodnou výměnou rovnic, tj. nalezením alespoň jednoho prvku $a_{i1} \neq 0$.

Uvedený postup lze vhodně zapsat pomocí maticových operací aplikovaných na matici

$$(A \mid \boldsymbol{b}) = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ a_{21} & \cdots & a_{2n} & b_2 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & b_n \end{array}\right).$$

První krok GEM vede na matici $\left(A' \mid \boldsymbol{b}'\right)$ tvaru

$$\left(A' \mid \boldsymbol{b}'\right) = \left(\begin{array}{cccc|c} a'_{11} & a'_{12} & \cdots & a'_{1n} & b'_1 \\ 0 & a'_{22} & \cdots & a'_{2n} & b'_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & a'_{n2} & \cdots & a'_{nn} & b'_n \end{array}\right)$$

a tento krok můžeme formálně popsat takto:

(a) Urči prvek $a_{r1} \neq 0$, $r = 1, 2, \ldots, n$, a pokračuj krokem (b); jestliže žádné takové r neexistuje, nelze pokračovat.

(b) Vyměň první a r-tý řádek matice $(A \mid \boldsymbol{b})$. Výsledkem je matice $(\bar{A} \mid \bar{\boldsymbol{b}})$.

(c) Pro $i = 2, 3, \ldots, n$, odečti násobek

$$l_{i1} = \frac{\bar{a}_{i1}}{\bar{a}_{11}}$$

prvního řádku od i-tého řádku matice $(\bar{A} \mid \bar{\boldsymbol{b}})$. Výsledkem je matice $(A' \mid \boldsymbol{b}')$. (4.6)

Čísla l_{i1} se nazývají *multiplikátory*. Prvek a_{r1} se nazývá *hlavním* prvkem nebo také *pivotem*.

Zapišme tento postup užitím maticového násobení:

$$(\bar{A} \mid \bar{\boldsymbol{b}}) = P_1(A \mid \boldsymbol{b}), \qquad (A' \mid \boldsymbol{b}') = G_1(\bar{A}, \bar{\boldsymbol{b}}) = G_1 P_1 \,(A, \boldsymbol{b})\,, \tag{4.7}$$

kde P_1 je *permutační* matice a G_1 je dolní trojúhelníková matice ($P_1, G_1 \in \mathcal{M}_n$):

$$P_1 = \begin{pmatrix} 0 & \cdots & & 1 & & & \\ \vdots & 1 & & \vdots & & & \\ & & \ddots & & & & \\ 1 & \cdots & & 0 & \cdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \qquad G_1 = \begin{pmatrix} 1 & \cdots & \cdots & 0 \\ -l_{21} & 1 & & \\ \vdots & & \ddots & \vdots \\ -l_{n1} & 0 & \cdots & 1 \end{pmatrix}.$$

Při násobení permutační maticí P_1 se vymění první a r-tý řádek matice A. Prvky permutační matice P_1 jsou dány vztahy $p_{ii} = 1$, $i \neq 1, r$, $p_{1r} = 1$, $p_{r1} = 1$, $p_{11} = 0$, $p_{rr} = 0$, $p_{ij} = 0$ pro ostatní i, j.

Matice G_1 se nazývá *Frobeniova* matice. Matice P_1, G_1 jsou regulární a platí

$$P_1^{-1} = P_1, \qquad G_1^{-1} = \begin{pmatrix} 1 & & & 0 \\ l_{21} & \ddots & & \\ \vdots & & \ddots & \\ l_{n1} & 0 & & 1 \end{pmatrix}.$$

Je zřejmé, že systémy $A\boldsymbol{x} = \boldsymbol{b}$, $A'\boldsymbol{x} = \boldsymbol{b}'$ mají stejné řešení. Je totiž:

$$A\boldsymbol{x}^* = \boldsymbol{b} \quad \Rightarrow \quad G_1P_1A\boldsymbol{x}^* = A'\boldsymbol{x}^* = \boldsymbol{b}' = G_1P_1\boldsymbol{b}$$

a

$$A'\boldsymbol{x} = \boldsymbol{b}' \quad \Rightarrow \quad P_1^{-1}G_1^{-1}A'\boldsymbol{x} = A\boldsymbol{x} = \boldsymbol{b} = P_1^{-1}G_1^{-1}\boldsymbol{b}' \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{x}^*.$$

Po prvním eliminačním kroku je výsledná matice $\left(A' \mid \boldsymbol{b}'\right)$ tvaru

$$(A' \mid \boldsymbol{b}') = \left(\begin{array}{cc|c} a'_{11} & \boldsymbol{a}'^T & \boldsymbol{b}'_1 \\ 0 & \tilde{A} & \tilde{\boldsymbol{b}} \end{array} \right),$$

kde $\tilde{A}$ je čtvercová matice řádu $n-1$. Nyní aplikujeme výše uvedený algoritmus (4.6) na systém $(\tilde{A} \mid \tilde{\boldsymbol{b}})$ a postup pak opět opakujeme. Označíme-li $(A^{(1)} \mid \boldsymbol{b}^{(1)}) = (A \mid \boldsymbol{b})$, $(A' \mid \boldsymbol{b}') = (A^{(2)} \mid \boldsymbol{b}^{(2)})$ atd., lze uvedenou proceduru zapsat takto:

$$(A \mid \boldsymbol{b}) = (A^{(1)} \mid \boldsymbol{b}^{(1)}) \to (A^{(2)} \mid \boldsymbol{b}^{(2)}) \to \ldots \to (A^{(n)} \mid \boldsymbol{b}^{(n)}) = (R \mid \boldsymbol{c}), \qquad (4.8)$$

kde R je požadovaná horní trojúhelníková matice. Matice $(A^{(k)} \mid \boldsymbol{b}^{(k)})$ v této

posloupnosti je tvaru

$$(A^{(k)} \mid \boldsymbol{b}^{(k)}) = \left(\begin{array}{cccccccc|c} \times & \cdots & \cdots & \times & \times & \cdots & \cdots & \times & \times \\ 0 & \times & \cdots & \times & \times & \cdots & \cdots & \times & \times \\ \vdots & & \ddots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & \cdots & \cdots & \times & \times & \cdots & \cdots & \times & \times \\ 0 & \cdots & \cdots & 0 & \times & \cdots & \cdots & \times & \times \\ \vdots & & & \vdots & \vdots & & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots & & & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & \times & \cdots & \cdots & \times & \times \end{array}\right) =$$

$$= \begin{pmatrix} A_{11}^{(k)} & A_{12}^{(k)} & \boldsymbol{b}_1^{(k)} \\ O & A_{22}^{(k)} & \boldsymbol{b}_2^{(k)} \end{pmatrix}. \tag{4.9}$$

Matice $A_{11}^{(k)}$ je horní trojúhelníková matice řádu $(k-1)$, $(k \geq 2)$. Přechod $(A^{(k)} \mid \boldsymbol{b}^{(k)}) \to (A^{(k+1)} \mid \boldsymbol{b}^{(k+1)})$ spočívá v aplikaci algoritmu (4.6) na matici $(A_{22}^{(k)} \mid \boldsymbol{b}_2^{(k)})$, což je matice typu $(n-k+1) \times (n-k+2)$. Prvky matic $A_{11}^{(k)}$, $A_{12}^{(k)}$ a vektoru $\boldsymbol{b}_1^{(k)}$ se při této transformaci nemění. Stejně jako v prvním kroku lze tuto transformaci vyjádřit maticově

$$(A^{(k)} \mid \boldsymbol{b}^{(k)}) = G_k P_k (A^{(k-1)} \mid \boldsymbol{b}^{(k-1)}) \tag{4.10}$$

$$(R, \boldsymbol{c}) = G_{n-1} P_{n-1} \ldots G_1 P_1 (A \mid \boldsymbol{b}) \tag{4.11}$$

s odpovídajícími permutačními maticemi P_k a Frobeniovými maticemi G_k tvaru

$$G_k = \begin{pmatrix} 1 & & \cdots & & 0 \\ & \ddots & & & \\ \vdots & & 1 & & \\ & & -l_{k+1,k} & \ddots & \\ & & \vdots & & \ddots \\ 0 & & -l_{n,k} & & 1 \end{pmatrix}.$$

Zde opět čísla l_{ik}, $i = k+1, \ldots, n$ se nazývají *multiplikátory.*

GEM rovněž dává velmi důležitý výsledek:

Věta 4.4. *Jestliže GEM lze provést bez výměny řádků, pak matici A lze rozložit na součin dolní a horní trojúhelníkové matice*

$$A = LR, \tag{4.12}$$

kde matice $R = (r_{ij})$, $L = (l_{ij})$ *jsou definovány takto:*

$$r_{ij} = \begin{cases} a_{ij}^{(i)}, & i = 1, \ldots, j \\ 0, & i = j+1, j+2, \ldots, n \end{cases} \tag{4.13}$$

a

$$l_{ij} = \begin{cases} 0, & i = 1, 2, \ldots, j-1,\ j \geq 2 \\ 1, & i = j,\ j = 1, \ldots, n \\ a_{ij}^{(j)}/a_{ii}^{(j)}, & i = j+1, \ldots, n, \end{cases} \tag{4.14}$$

l_{ij} jsou příslušné multiplikátory dané algoritmem (4.6).

Důkaz. Jestliže neměníme pořadí řádků, je $P_1 = P_2 = \ldots = P_{n-1} = E$. Nyní z (4.10) a (4.11) plyne, že

$$R = G_{n-1} \ldots G_1 A,$$

a tedy

$$G_1^{-1} G_2^{-1} \ldots G_{n-1}^{-1} R = A. \tag{4.15}$$

Dále

$$G_j^{-1} = \begin{pmatrix} 1 & & \cdots & & 0 \\ & \ddots & & & \\ \vdots & & 1 & & \\ & & l_{j+1,j} & \ddots & \\ & & \vdots & & \ddots \\ 0 & & l_{nj} & & 1 \end{pmatrix}.$$

Odtud je zřejmé, že

$$G_1^{-1} \ldots G_{n-1}^{-1} = \begin{pmatrix} 1 & \cdots & \cdots & 0 \\ l_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{pmatrix}.$$

Z algoritmu (4.6) plyne, že prvky matice R jsou dány vztahy (4.13). Položíme-li nyní $L = G_1^{-1} \ldots G_{n-1}^{-1}$, plyne ze (4.15) tvrzení věty. □

Poznámka 2. Jestliže GEM nelze provést bez výměny řádků, definuje tento algoritmus rozklad matice PA:

$$PA = LR, \tag{4.16}$$

kde $P = P_{n-1} P_{n-2} \ldots P_1$.

Při praktickém výpočtu ovšem nevíme předem, které řádky budeme muset vyměnit, takže není možné určit součin PA a pak provést rozklad. Rozklad je ovšem možné provést pomocí následující úvahy:
Nechť například je třeba při výpočtu vyměnit k-tý a j-tý řádek, $k < j$. Pokud bychom to věděli předem a provedli výměnu před začátkem vlastní GEM, pak v matici $A^{(j)}$, by byly oproti současnému stavu vyměněny celé řádky k-tý a j-tý, v matici obsahující mezivýsledek pro výpočet L by byly vyměněny jen spočítané části obou řádků, tedy sloupce $1, \ldots, k-1$. Můžeme tedy provést příslušné výměny a zaznamenat si, které řádky byly vyměněny.

Pro toto zaznamenání není potřeba pracovat s celými permutačními maticemi, ale stačí tzv. permutační vektor $\boldsymbol{p}$. Na začátku výpočtu položíme $\boldsymbol{p} = (1, \ldots, n)^T$ a při výměně řádků vyměníme stejné řádky i ve vektoru $\boldsymbol{p}$. Pak jeho i-tá složka p_i udává původní číslo řádku matice A v matici PA. Tj. jestliže například po ukončení výpočtu je $\boldsymbol{p} = (3, 5, 1, 4, 2)^T$, pak matice PA v (4.16) je postupně tvořena třetím, pátým, prvním, čtvrtým a druhým řádkem matice A.

Trojúhelníkový rozklad (4.12) resp. (4.16) má velký význam pro řešení systémů lineárních rovnic. Jestliže známe rozklad (4.16), pak systém $A\boldsymbol{x} = \boldsymbol{b}$ lze ihned řešit pro libovolný vektor $\boldsymbol{b}$. Je totiž

$$PA\boldsymbol{x} = LR\boldsymbol{x} = P\boldsymbol{b}.$$

Řešení $\boldsymbol{x}^*$ nyní najdeme řešením dvou systémů s trojúhelníkovými maticemi:

$$L\boldsymbol{u} = P\boldsymbol{b}, \qquad R\boldsymbol{x} = \boldsymbol{u}$$

za předpokladu $r_{ii} \neq 0$, $i = 1, \ldots, n$. První systém má dolní trojúhelníkovou matici a řešíme jej tedy od první rovnice, druhý systém má horní trojúhelníkovou matici a řešíme jej od poslední rovnice.

Podívejme se na počet násobení a dělení pro přímý chod GEM.

Podle algoritmu (4.6) je v prvním kroku zapotřebí vypočítat $(n-1)$ multiplikátorů, což znamená $(n-1)$ dělení. Každý prvek první rovnice včetně pravé strany musí být násoben každým multiplikátorem, tzn. že v prvním kroku je zapotřebí celkem $n - 1 + n(n-1) = (n-1)(n+1)$ násobení a dělení (násobicích operací). Na j-tém kroku se pak požaduje $(n-j) + (n-j+1)(n-j)$ násobicích operací.

Připomeňme nyní, že

$$\sum_{j=1}^{m} 1 = m, \qquad \sum_{j=1}^{m} j = \frac{m(m+1)}{2}, \qquad \sum_{j=1}^{m} j^2 = \frac{m(m+1)(2m+1)}{6}.$$

Pak celkový počet násobicích operací pro přímý chod je roven

$$\sum_{j=1}^{n-1} (n-j)(n-j+2) = \frac{2n^3 + 3n^2 - 5n}{6}.$$

Podobně lze ukázat, že pro zpětný chod, tj. pro řešení systému $R\boldsymbol{x} = \boldsymbol{c}$ je zapotřebí

$$1 + \sum_{j=1}^{n-1} ((n-j)+1) = \frac{n^2 + n}{2}$$

násobicích operací. Celkový počet násobicích operací pro GEM je roven

$$\frac{2n^3 + 3n^2 - 5n}{6} + \frac{n^2 + n}{2} = \frac{n^3 + 3n^2 - n}{3} \approx \frac{n^3}{3}.$$

Obdobným způsobem lze spočítat počet sčítání a odčítání. Počet těchto operací pro GEM je:

$$\frac{2n^3 + 3n^2 - 5n}{6} \approx \frac{n^3}{3}.$$

Je zřejmé, že počet aritmetických operací velmi rychle roste s rostoucím n. Tento fakt ukazuje pro některá n následující tabulka ([4]):

n	násobení/dělení	sčítání/odčítání
3	17	11
5	65	50
10	430	375
50	44150	42875
100	343300	338250

Zmíníme se nyní o problémech s výběrem pivotů. Viděli jsme, že GEM selhává, jestliže hlavní prvek je roven nule. V tomto případě lze vyměnit pořadí rovnic. Ale problematická situace nastává, jestliže některý z pivotů je blízký nule: v tomto případě lze výpočet provést, ale získané výsledky mohou být zcela chybné. Ilustrujme tuto skutečnost na známém příkladu Forsytha a Molera (viz [6]):

Příklad 4.1. Aplikujme GEM na matici

$$A = \begin{pmatrix} 0,0001 & 1 \\ 1 & 1 \end{pmatrix}.$$

Multiplikátor $l_{21} = 1/10^{-4} = 10^4$. Matice

$$G = \begin{pmatrix} 1 & 0 \\ -10^4 & 1 \end{pmatrix}$$

a

$$A_2^{(2)} = R = \begin{pmatrix} 0,0001 & 1 \\ 0 & -10^4 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 \\ 10^4 & 1 \end{pmatrix},$$

neboť $1 - 10^4 \approx -10^4$. GEM definuje rozklad matice A na součin matic L a R, což je ale v tomto případě

$$L\,R = \begin{pmatrix} 0,0001 & 1 \\ 1 & 0 \end{pmatrix},$$

tedy tato matice se nerovná matici A. Výsledek lze vysvětlit faktem, že prvek $a_{11}^{(1)} = 0,0001$ je velmi malý, což má za následek velký multiplikátor a při jeho použití se prakticky vyloučí malé vstupní prvky ($1 - 10^4 \approx -10^4$). Tomuto problému se můžeme vyhnout výměnou řádků matice A.

Uvažujme tedy matici

$$P_1 A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0,0001 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0,0001 & 1 \end{pmatrix}$$

a aplikujeme GEM na matici P_1A. Nyní $l_{21} = 10^{-4}$.

$$R = G_1P_1A = \begin{pmatrix} 1 & 0 \\ -10^{-4} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -10^{-4} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

$(1 - 10^{-4} \approx 1)$,

$$L = \begin{pmatrix} 1 & 0 \\ -10^{-4} & 1 \end{pmatrix} \Rightarrow LR = \begin{pmatrix} 1 & 1 \\ -10^{-4} & 1 \end{pmatrix} = P_1A.$$

Tomuto postupu obecně říkáme Gaussova eliminace s *částečným výběrem hlavního prvku (pivota)*. Tento postup spočívá v tom, že v každém kroku vybíráme v příslušném sloupci prvek maximální v absolutní hodnotě, tj. určíme p tak, aby

$$|a_{pk}^{(k)}| = \max_{k \le i \le n} |a_{ik}^{(k)}|$$

a vyměníme p-tou a k-tou rovnici.

Další vhodnou modifikací GEM je *úplný výběr pivota*. Tato procedura znamená, že na j-tém kroku vybíráme prvek maximální v absolutní hodnotě ze submatice $A_{22}^{(k)}$, tj.

$$\left|a_{rs}^{(k)}\right| = \max_{i,j=k,\ldots,n} \left|a_{ij}^{(k)}\right|.$$

Pak vyloučíme neznámou x_s pomocí r-té rovnice ze zbývajících $(n-j)$ rovnic. Pro provedení 1. kroku této procedury je třeba $n^2 - 1$ porovnání absolutních hodnot koeficientů. Druhý krok vyžaduje $(n-1)^2 - 1$ porovnání a celkový počet porovnání je roven

$$\sum_{k=2}^{n} (k^2 - 1) = \frac{n(n-1)(2n+5)}{6}.$$

Poznámka 3. Výměnu řádků u GEM s částečným výběrem pivota resp. řádků a sloupců u GEM s úplným výběrem pivota lze realizovat opět prostřednictvím permutačních matic.

Příklad 4.2. Systém

$$\begin{aligned} 2x_1 + 4x_2 - x_3 &= -5 \\ x_1 + x_2 - 3x_3 &= -9 \\ 4x_1 + x_2 + 2x_3 &= 9 \end{aligned}$$

řešte

1) GEM bez výběru pivota,

2) GEM s částečným výběrem pivota.

Řešení.

1) GEM bez výběru pivota:

$$(A \mid \boldsymbol{b}) = (A^{(1)} \mid \boldsymbol{b}^{(1)}) = \left(\begin{array}{ccc|c} 2 & 4 & -1 & -5 \\ 1 & 1 & -3 & -9 \\ 4 & 1 & 2 & 9 \end{array}\right)$$

Prvek $a_{11} = 2 \;\Rightarrow\; l_{21} = \frac{1}{2}$, $l_{31} = 2$.

$$(A^{(2)} \mid \boldsymbol{b}^{(2)}) = \left(\begin{array}{ccc|c} 2 & 4 & -1 & -5 \\ 0 & -1 & -\frac{5}{2} & -\frac{13}{2} \\ 0 & -7 & 4 & 19 \end{array}\right)$$

Prvek $a_{22}^{(2)} = -1 \;\Rightarrow\; l_{31} = 7$.

$$(A^{(3)} \mid \boldsymbol{b}^{(3)}) = \left(\begin{array}{ccc|c} 2 & 4 & -1 & -5 \\ 0 & -1 & -\frac{5}{2} & -\frac{13}{2} \\ 0 & 0 & \frac{43}{2} & \frac{129}{2} \end{array}\right)$$

Nyní řešíme systém $A^{(3)}\boldsymbol{x} = \boldsymbol{b}^{(3)}$, tj.

$$\begin{aligned} 2x_1 + 4x_2 - x_3 &= -5 \\ -x_2 - \frac{5}{2}x_3 &= -\frac{13}{2} \\ \frac{43}{2}x_3 &= \frac{129}{2} \end{aligned}$$

Řešíme od poslední rovnice (zpětný chod):

$$x_3 = 3, \quad x_2 = -1, \quad x_1 = 1.$$

Matice L je v tomto případě tvaru:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 2 & 7 & 1 \end{pmatrix}$$

a matice $R = A^{(3)}$

$$R = \begin{pmatrix} 2 & 4 & -1 \\ 0 & -1 & -\frac{5}{2} \\ 0 & 0 & \frac{43}{2} \end{pmatrix}.$$

Snadno se ověří, že $A = LR$.

2) GEM s částečným výběrem pivota:

$$(A \mid \boldsymbol{b}) = (A^{(1)} \mid \boldsymbol{b}^{(1)}) = \left(\begin{array}{ccc|c} 2 & 4 & -1 & -5 \\ 1 & 1 & -3 & -9 \\ 4 & 1 & 2 & 9 \end{array}\right)$$

Je $|a_{31}| = \max\limits_{1 \le i \le 3} |a_{i1}| \Rightarrow$ vyměníme 1. a 3. rovnici:

$$(\bar{A} \mid \bar{\boldsymbol{b}}) = \left(\begin{array}{ccc|c} 4 & 1 & 2 & 9 \\ 1 & 1 & -3 & -9 \\ 2 & 4 & -1 & -5 \end{array}\right);$$

$$\text{permutační matice } P_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Hlavní prvek $\bar{a}_{11} = 4 \Rightarrow l_{21} = \frac{1}{4}$, $l_{31} = \frac{1}{2}$. Odtud

$$(A^{(2)} \mid \boldsymbol{b}^{(2)}) = \left(\begin{array}{ccc|c} 4 & 1 & 2 & 9 \\ 0 & \frac{3}{4} & -\frac{7}{2} & -\frac{45}{4} \\ 0 & \frac{7}{2} & -2 & -\frac{19}{2} \end{array}\right).$$

Jelikož $|a_{32}^{(2)}| = \max(|a_{22}^{(2)}|, |a_{32}^{(2)}|)$, vyměníme 2. a 3. řádek, tj. matici $A^{(2)}$ vynásobíme permutační maticí P_2, která je tvaru

$$P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

tj.

$$P_2(A^{(2)} \mid \boldsymbol{b}^{(2)}) = \left(\begin{array}{ccc|c} 4 & 1 & 2 & 9 \\ 0 & \frac{7}{2} & -2 & -\frac{19}{2} \\ 0 & \frac{3}{4} & -\frac{7}{2} & -\frac{45}{4} \end{array}\right).$$

Odpovídající multiplikátor $l_{32} = \frac{3}{14}$ a výsledná matice

$$(A^{(3)} \mid \boldsymbol{b}^{(3)}) = \left(\begin{array}{ccc|c} 4 & 1 & 2 & 9 \\ 0 & \frac{7}{2} & -2 & -\frac{19}{2} \\ 0 & 0 & -\frac{43}{14} & -\frac{129}{14} \end{array}\right)$$

Tento systém opět řešíme od poslední rovnice a opět dostaneme:

$$x_3 = 3, \quad x_2 = -1, \quad x_1 = 1.$$

Dále

$$P_2P_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$P_2P_1A = \begin{pmatrix} 4 & 1 & 2 \\ 2 & 4 & -1 \\ 1 & 1 & -3 \end{pmatrix}.$$

Snadno se ověří, že GEM definuje tento rozklad matice PA, $P = P_2P_1$:

$$PA = LR, \qquad \text{kde } L = P_2P_1(G_2P_2G_1P_1)^{-1},$$

a tedy

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{4} & \frac{3}{14} & 1 \end{pmatrix}, \qquad R = \begin{pmatrix} 4 & 1 & 2 \\ 0 & \frac{7}{2} & -2 \\ 0 & 0 & -\frac{43}{14} \end{pmatrix}.$$

Nyní se ještě zmíníme o aplikaci GEM na speciální typy matic (důkazy viz [4], [13]).

Věta 4.5.

a) *Nechť matice A je ryze řádkově diagonálně dominantní. Pak GEM lze provést bez výměny řádků a sloupců.*

b) *Nechť matice A je pozitivně definitní. Pak GEM lze provést bez výměny řádků a sloupců.*

Za předpokladu, že všechny pivoty $a_{kk}^{(k)}$, $k = 1, \ldots, n$, jsou různé od nuly, GEM definuje rozklad matice A na součin dolní a horní trojúhelníkové matice $A = LR$ (říkáme, že se jedná o LR rozklad nebo LR faktorizaci). Z algoritmu GEM je zřejmé, že tento přímý rozklad nemusí existovat dokonce i pro velmi jednoduché matice. Zabývejme se nyní otázkou, kdy lze takový přímý rozklad provést.

Věta 4.6. *Nechť všechny hlavní minory matice $A \in \mathcal{M}_n$ jsou různé od nuly, tj.*

$$a_{11} \neq 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \qquad \ldots \qquad , \det A \neq 0.$$

Pak matici A lze rozložit na součin dolní a horní trojúhelníkové matice.

Důkaz. (indukcí) Je-li $n = 1$ a prvek $a_{11} \neq 0$, je jasné, že tvrzení věty platí, neboť existují nenulová čísla c_{11}, b_{11} tak, že $a_{11} = c_{11}b_{11}$, tj. $A = CB$, $C = (c_{11})$, $B = (b_{11})$.

Nechť nyní podle indukčního předpokladu tvrzení platí pro matici $A_{n-1} \in \mathcal{M}_{n-1}$ a dokážeme, že platí i pro matici $A \in \mathcal{M}_n$.

Matici A_{n-1} lze tedy vyjádřit ve tvaru $A_{n-1} = C_{n-1}B_{n-1}$, kde C_{n-1} je dolní trojúhelníková matice a B_{n-1} je horní trojúhelníková matice. Matici A zapíšeme blokově takto:

$$A = \begin{pmatrix} & & & a_{1n} \\ & A_{n-1} & & a_{2n} \\ & & & \vdots \\ & & & a_{n-1,n} \\ a_{n1} & \cdots & a_{n,n-1} & a_{nn} \end{pmatrix} = \begin{pmatrix} A_{n-1} & \boldsymbol{u} \\ \boldsymbol{v}^T & a_{nn} \end{pmatrix},$$

kde vektory $\boldsymbol{u}, \boldsymbol{v}$ jsou vektory dimenze $n-1$. Hledejme nyní matice C, B, $A = CB$, rovněž v blokovém tvaru:

$$C = \begin{pmatrix} C_{n-1} & \boldsymbol{o} \\ \boldsymbol{x}^T & c_{nn} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{n-1} & \boldsymbol{y} \\ \boldsymbol{o}^T & b_{nn} \end{pmatrix}.$$

Zde $\boldsymbol{x}, \boldsymbol{y}$ jsou neznámé vektory řádu $n-1$ a b_{nn}, c_{nn} jsou neznámé prvky matic B, C. Podle pravidla o násobení blokově daných matic dostáváme z rovnice $A = CB$:

$$\begin{pmatrix} A_{n-1} & \boldsymbol{u} \\ \boldsymbol{v}^T & a_{nn} \end{pmatrix} = \begin{pmatrix} C_{n-1}B_{n-1} & C_{n-1}\boldsymbol{y} \\ \boldsymbol{x}^T B_{n-1} & \boldsymbol{x}^T\boldsymbol{y} + c_{nn}b_{nn} \end{pmatrix}. \tag{4.17}$$

Podle předpokladu je matice A_{n-1} regulární a tedy jsou regulární i matice C_{n-1}, B_{n-1}. Porovnáme-li ve vztahu (4.17) prvky v odpovídajících pozicích, dostaneme:

$$\begin{aligned} A_{n-1} &= C_{n-1}B_{n-1} \\ C_{n-1}\boldsymbol{y} &= \boldsymbol{u} \\ \boldsymbol{x}^T B_{n-1} &= \boldsymbol{v}^T \\ \boldsymbol{x}^T\boldsymbol{y} + c_{nn}b_{nn} &= a_{nn} \end{aligned} \tag{4.18}$$

Ze vztahů (4.18) lze určit vektory $\boldsymbol{x}, \boldsymbol{y}$ a čísla c_{nn}, b_{nn}, z nichž jedno lze volit libovolně ($\neq 0$). To znamená, že existuje dolní trojúhelníková matice C a horní trojúhelníková matice B tak, že $A = CB$. □

Poznámka 4. Předepíšeme-li matici C diagonální prvky rovny 1, je rozklad jednoznačný. Na druhé straně, GEM bez výběru pivota rovněž definuje rozklad $A = LR$. Odtud plyne, že $L = C$, $R = B$. Odtud plyne, že neexistence přímého rozkladu matice a selhání GEM bez výběru pivota se dá objasnit stejnými příčinami.

Příklad 4.3. Je dána matice A:

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

Pokusme se matici A rozložit na součin CB. Porovnáním prvků v pozici (i,j) matice A a prvků v pozici (i,j) matice CB dostaneme (předpokládáme $c_{ii}=1$, $i=1,\dots,n$)

$$\begin{array}{l|l|cl} 1=b_{11} & 2=c_{21}b_{11} & \Rightarrow & c_{21}=2 \\ 2=b_{12} & 4=c_{21}b_{12}+b_{22} & \Rightarrow & b_{22}=0 \\ 3=b_{13} & 1=c_{21}b_{13}+b_{23} & \Rightarrow & b_{23}=-5 \end{array}$$

$$\begin{array}{ll} 4=c_{31}b_{11} & \Rightarrow \quad c_{31}=4 \\ 5=c_{31}b_{12}+c_{32}b_{22} & \Rightarrow \quad \text{nelze určit } c_{32}, \text{ neboť } b_{22}=0. \\ 2=c_{31}b_{13}+c_{32}b_{23}+b_{33} & \end{array}$$

Odtud

$$C=\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & ? & . \end{pmatrix} \qquad B=\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -5 \\ . & . & . \end{pmatrix}$$

Rozklad není možný, neboť $\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix}=0$.

V příkladě 4.2 v části 1) byl proveden přímý rozklad matice na součin horní a dolní trojúhelníkové matice. Pokud bychom chtěli udělat rozklad

$$\begin{pmatrix} 2 & 4 & -1 \\ 1 & 1 & -3 \\ 4 & 1 & 2 \end{pmatrix}=\begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix}=\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

porovnáním jednotlivých prvků, dostaneme soustavu rovnic:

$$\begin{array}{rcrrcrrcl} l_{11}u_{11} & = & 2 & l_{21}u_{11} & = & 1 & l_{31}u_{11} & = & 4 \\ l_{11}u_{12} & = & 4 & l_{21}u_{12}+l_{22}u_{22} & = & 1 & l_{31}u_{12}+l_{32}u_{22} & = & 1 \\ l_{11}u_{13} & = & -1 & l_{21}u_{13}+l_{22}u_{23} & = & -3 & l_{31}u_{13}+l_{32}u_{23}+l_{33}u_{33} & = & 2 \end{array}$$

Máme 9 rovnic pro 12 neznámých, řešení tedy není určeno jednoznačně. Rozklad vypočtený v příkladě 4.2 bychom dostali volbou $l_{11}=l_{22}=l_{33}=1$.

I když GEM definuje rozklad matice na součin trojúhelníkových matic, je někdy vhodné mít k dispozici algoritmus, který tuto faktorizaci provede přímo. Tento postup je vhodný zejména v případech, kdy je třeba řešit více systémů s toutéž maticí A, (viz [4], [13]).

§ 4.3. Systémy se speciálními maticemi

Předchozí věty lze s výhodou užít i v případech, kdy matice A má speciální strukturu a předpokládat speciální tvar rozkladu matice.

Věta 4.7. *Nechť matice $A \in \mathcal{M}_n$ je symetrická a splňuje předpoklady předchozí věty. Pak existuje taková horní trojúhelníková matice $T \in \mathcal{M}_n$, že $A = T^T T$.*

Důkaz. Podle předchozí věty existuje rozklad $A = CB$. Nechť $c_{ii} = 1, i = 1, \ldots, n$ a rozklad je tedy jednoznačný. Nechť D je diagonální matice s prvky $d_{ii} = b_{ii}$, $i = 1, \ldots, n$, na diagonále. Položme $\bar{B} = D^{-1}B$. Pak A lze zapsat ve tvaru

$$A = CD\bar{B}.$$

Zde $\bar{B}$ je horní trojúhelníková matice s jedničkami na diagonále. Dále platí

$$A = A^T = \left(CD\bar{B}\right)^T = \bar{B}^T DC^T.$$

$\bar{B}^T$ je dolní trojúhelníková matice s jedničkami na diagonále, DC^T je horní trojúhelníková matice. Z jednoznačnosti rozkladu plyne:

$$\bar{B}^T = C, \qquad DC^T = B.$$

Nyní položíme

$$T = \sqrt{D}\bar{B},$$

kde $\sqrt{D}$ je diagonální matice s prvky $\sqrt{d_{ii}} = \sqrt{b_{ii}}$, $i = 1, \ldots, n$, na diagonále.

Nyní

$$T^T T = \bar{B}^T \sqrt{D}\sqrt{D}\bar{B} = \bar{B}^T D\bar{B} = A.$$

Dostali jsme požadovaný rozklad matice A. □

Důsledek. *Nechť T je matice uvedená v předchozí větě. Prvky této matice jsou určeny vztahy:*

$$\begin{aligned}
t_{11} &= \sqrt{a_{11}} \\
t_{1j} &= \frac{a_{1j}}{t_{11}}, && j = 2, \ldots, n \\
t_{ii} &= \sqrt{a_{ii} - \sum_{l=1}^{i-1} t_{li}^2}, && i = 2, \ldots, n \\
t_{ij} &= \frac{1}{t_{ii}} \left(a_{ij} - \sum_{l=1}^{i-1} t_{li} t_{lj} \right) && \text{pro } j > i \\
t_{ij} &= 0 && \text{pro } i > j.
\end{aligned} \tag{4.19}$$

Důkaz plyne ihned porovnáním odpovídajících prvků ve vztahu $A = T^T T$.

Uvedená metoda se nazývá metoda *Choleského* nebo také metoda druhých odmocnin.

Poznámka 5. Je-li A pozitivně definitní matice, probíhá výpočet bez komplikací. V tomto případě jsou všechny prvky matice T reálné. Obecně může mít matice T ryze imaginární prvky. Tyto prvky se vyskytují v celém řádku matice a při dalším

výpočtu se imaginární jednotky vyruší. Tento rozklad vede opět na řešení dvou systémů s trojúhelníkovými maticemi:

$$T^T \boldsymbol{z} = \boldsymbol{b}, \qquad T\boldsymbol{x} = \boldsymbol{z}.$$

Počet násobicích operací Choleského metody je přibližně $n^3/6$; přitom je třeba ještě vyčíslit n druhých odmocnin.

Příklad 4.4. Choleského metodou řešte systém

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 1 \\ 2x_1 + 2x_2 + 4x_3 &= 3 \\ -x_1 + 4x_2 + 8x_3 &= 6. \end{aligned}$$

Řešení. Najdeme rozklad matice A ve tvaru $T^T T = A$. Prvky matice T vypočteme ze vztahů (4.19).

$$\begin{aligned} &t_{11} = 1, \quad t_{12} = 2, \quad t_{13} = -1, \\ &t_{22} = \sqrt{a_{22} - t_{12}^2} = i\sqrt{2}, \\ &t_{23} = \frac{1}{t_{22}}(a_{23} - t_{13}t_{12}) = -i3\sqrt{2}, \\ &t_{33} = \sqrt{a_{33} - (t_{13}^2 + t_{23}^2)} = 5. \end{aligned}$$

Matice je tvaru

$$T = \begin{pmatrix} 1 & 2 & -1 \\ 0 & i\sqrt{2} & -i3\sqrt{2} \\ 0 & 0 & 5 \end{pmatrix}.$$

Nyní řešíme $T^T \boldsymbol{z} = \boldsymbol{b}$, tj.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & i\sqrt{2} & 0 \\ -1 & -i3\sqrt{2} & 5 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 6 \end{pmatrix}.$$

Řešení tohoto systému je vektor $\boldsymbol{z} = (1, -i\sqrt{2}/2, 2)^T$. Nyní řešíme $T\boldsymbol{x} = \boldsymbol{z}$

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & i\sqrt{2} & -i3\sqrt{2} \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -i\sqrt{2}/2 \\ 2 \end{pmatrix}.$$

Řešení tohoto systému (a tedy i řešení daného systému) je vektor

$$\boldsymbol{x} = (0, \tfrac{7}{10}, \tfrac{2}{5})^T.$$

Přímý rozklad matice na součin trojúhelníkových matic lze také použít pro

třídiagonální matice. Uvažujme třídiagonální matici A:

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & a_{n,n-1} & a_{nn} \end{pmatrix}$$

Hledejme rozklad matice A ve tvaru: $A = LU$

$$L = \begin{pmatrix} l_{11} & 0 & & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ 0 & l_{32} & l_{33} & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & l_{n,n-1} & l_{nn} \end{pmatrix},$$

$$U = \begin{pmatrix} 1 & u_{12} & 0 & \cdots & 0 \\ 0 & 1 & u_{23} & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & 1 & u_{n-1,n} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \tag{4.20}$$

Je třeba určit $(2n-1)$ prvků matice L a $(n-1)$ prvků matice U, tedy celkem $(3n-2)$ prvků. Tyto prvky lze určit z následujících rovnic:

$$\begin{array}{ll} a_{11} = l_{11} & \\ a_{i,i-1} = l_{i,i-1}, & i = 2,3,\ldots,n \\ a_{ii} = l_{i,i-1}u_{i-1,i} + l_{ii}, & i = 2,3,\ldots,n \\ a_{i,i+1} = l_{ii}u_{i,i+1}, & i = 1,2,\ldots,n-1. \end{array} \tag{4.21}$$

Tyto rovnice se snadno získají porovnáním prvků matice A s odpovídajícími prvky součinu LU. Uvedená metoda se nazývá *Croutova*.

Věta 4.8. *Nechť $A \in \mathcal{M}_n$ je třídiagonální matice s vlastnostmi:*

$$\left.\begin{array}{ll} a_{i,i-1}a_{i,i+1} \neq 0, & i = 2,3,\ldots,n-1, \\ |a_{11}| > |a_{12}|, & \\ |a_{ii}| \geq |a_{i,i-1}| + |a_{i,i+1}|, & i = 2,\ldots,n-1, \\ |a_{nn}| > |a_{n,n-1}|. & \end{array}\right\} \begin{array}{l} A \text{ řádkově diagonálně} \\ \text{dominantní} \end{array}$$

Pak matice A je regulární a hodnoty l_{ii}, $i = 1,\ldots,n$, vypočtené ze vztahů (4.21) *jsou různé od nuly.*

Důkaz viz [4].

Důsledek. *Jsou-li splněny předpoklady věty* 4.8*, lze matici A rozložit na součin dolní a horní trojúhelníkové matice tvaru* (4.20).

Poznámka 6. Počet násobicích operací pro realizaci Croutovy metody je $(5n-4)$, počet sčítacích operací $(3n-3)$.

Jestliže matici A vyjádříme ve tvaru $A = LU$, pak systém $A\boldsymbol{x} = \boldsymbol{b}$ lze opět jednoduše řešit takto:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \Rightarrow \quad \left\{ \begin{array}{rcl} L\boldsymbol{y} & = & \boldsymbol{b} \\ U\boldsymbol{x} & = & \boldsymbol{y}. \end{array} \right.$$

Příklad 4.5. Croutovou metodou řešte systém

$$\begin{aligned} 2x_1 - x_2 &= 4 \\ x_1 + 4x_2 + x_3 &= 5 \\ x_2 + 3x_3 - 2x_4 &= -1 \\ 2x_3 - 3x_4 &= \tfrac{7}{5} \end{aligned}$$

Řešení. Podle vztahů (4.21) určíme prvky matic L a U:

$$\begin{aligned} i = 1 \qquad & l_{11} = a_{11} = 2, \quad u_{12} = \frac{a_{12}}{l_{11}} = -\tfrac{1}{2} \\ i = 2 \qquad & l_{22} = a_{22} - l_{21}u_{12} = 4 - 1(-\tfrac{1}{2}) = \tfrac{9}{2} \\ & l_{21} = a_{21} = 1 \\ & u_{23} = \frac{a_{23}}{l_{22}} = \frac{1}{\frac{9}{2}} = \tfrac{2}{9} \\ i = 3 \qquad & l_{32} = a_{32} = 1, \quad l_{33} = a_{33} - l_{32}u_{23} = 3 - 1.\tfrac{2}{9} = \tfrac{25}{9} \\ & u_{34} = \frac{a_{34}}{l_{33}} = -\frac{2}{\frac{25}{9}} = -\tfrac{18}{25} \\ i = 4 \qquad & l_{43} = a_{43} = 2, \quad l_{44} = a_{44} - l_{43}u_{34} = -3 - 2(-\tfrac{18}{25}) = -\tfrac{39}{25} \end{aligned}$$

Matice L a U jsou tvaru

$$L = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & \frac{9}{2} & 0 & 0 \\ 0 & 1 & \frac{25}{9} & 0 \\ 0 & 0 & 2 & -\frac{39}{25} \end{pmatrix}, \quad U = \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{2}{9} & 0 \\ 0 & 0 & 1 & -\frac{18}{25} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Nyní řešíme systém $L\boldsymbol{y} = \boldsymbol{b}$. Řešením je vektor $\boldsymbol{y} = (2, \frac{2}{3}, -\frac{3}{5}, -\frac{5}{3})^T$. Nyní řešíme systém $U\boldsymbol{x} = \boldsymbol{y}$. Řešením tohoto systému, a tedy i daného systému, je vektor $\boldsymbol{x} = (\frac{38}{15}, \frac{16}{15}, -\frac{9}{5}, -\frac{5}{3})^T$.

§ 4.4. Výpočet inverzní matice a determinantu

S problémem řešení systému $A\boldsymbol{x} = \boldsymbol{b}$ souvisí také problémy výpočtu inverzní matice a determinantu matice.

Výpočet inverzní matice k matici A je ekvivalentní řešení systému

$$AX = E,$$

kde $X = A^{-1}$, E je jednotková matice. Nechť $X = (x_{ij})$. Pak řešit systém $AX = E$ znamená řešit n systémů tvaru

$$A\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad A\begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots \quad A\begin{pmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

tj. řešíme n systémů s toutéž maticí A a s různými pravými stranami. K řešení těchto systémů lze užít některé z již uvedených přímých metod (např. GEM).

Příklad 4.6. Užitím GEM nalezněte matici inverzní k matici

$$A = \begin{pmatrix} 2 & -7 & 4 \\ 1 & 9 & -6 \\ -3 & 8 & 5 \end{pmatrix}.$$

Řešení. Řešit systém $AX = E$, kde $X = (x_{ij})$ je inverzní matice, znamená řešit n systémů vždy s toutéž maticí A. Postup při aplikaci GEM zapíšeme do tabulky:

a_{i1}	a_{i2}	a_{i3}	$\boldsymbol{b}^1$	$\boldsymbol{b}^2$	$\boldsymbol{b}^3$
2	-7	4	1	0	0
1	9	-6	0	1	0
-3	8	5	0	0	1
2	-7	4	1	0	0
0	$\frac{25}{2}$	-8	$-\frac{1}{2}$	1	0
0	$-\frac{5}{2}$	11	$\frac{3}{2}$	0	1
2	-7	4	1	0	0
0	$\frac{25}{2}$	-8	$-\frac{1}{2}$	1	0
0	0	$\frac{47}{5}$	$\frac{7}{5}$	$\frac{1}{5}$	1

Prvky inverzní matice $X = (x_{ij})$ získáme řešením systémů rovnic

$$\begin{pmatrix} 2 & -7 & 4 \\ 0 & \frac{25}{2} & -8 \\ 0 & 0 & \frac{47}{5} \end{pmatrix}\begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ -\frac{1}{2} \\ \frac{7}{5} \end{pmatrix},$$

$$\begin{pmatrix} 2 & -7 & 4 \\ 0 & \frac{25}{2} & -8 \\ 0 & 0 & \frac{47}{5} \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ \frac{1}{5} \end{pmatrix},$$

$$\begin{pmatrix} 2 & -7 & 4 \\ 0 & \frac{25}{2} & -8 \\ 0 & 0 & \frac{47}{5} \end{pmatrix} \begin{pmatrix} x_{13} \\ x_{23} \\ x_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Výsledná matice je tvaru

$$X = A^{-1} = \frac{1}{235} \begin{pmatrix} 93 & 67 & 6 \\ 13 & 22 & 16 \\ 35 & 5 & 25 \end{pmatrix}.$$

Je třeba poznamenat, že výpočet inverzní matice je třikrát „dražší“ než řešení systému $A\boldsymbol{x} = \boldsymbol{b}$. Z těchto důvodů je vhodné se „vyhnout“ přímému výpočtu A^{-1} kdykoliv je to možné. Lineární systém bychom nikdy neměli řešit explicitním výpočtem inverzní matice.

Pro zajímavost uvedeme ještě dva užitečné vzorce pro výpočet inverze matice B, která se poněkud liší od matice A ([6]):

1. *Shermanův-Morrisonův vzorec.* Nechť $\boldsymbol{u}$, $\boldsymbol{v}$ jsou vektory, $A \in \mathcal{M}_n$ je regulární matice. Pak
$$(A - \boldsymbol{u}\boldsymbol{v}^T)^{-1} = A^{-1} + \alpha(A^{-1}\boldsymbol{u}\boldsymbol{v}^T A^{-1}),$$
kde
$$\alpha = \frac{1}{(1 - \boldsymbol{v}^T A^{-1}\boldsymbol{u})},$$
za předpokladu $\boldsymbol{v}^T A^{-1}\boldsymbol{u} \neq 1$.

2. *Woodburyho vzorec.* Nechť $A, U, V \in \mathcal{M}_n$,
$$(A - UV^T)^{-1} = A^{-1} + A^{-1}U(E - V^T A^{-1}U)^{-1}V^T A^{-1},$$
za předpokladu, že $E - V^T A^{-1}U$ je regulární.

Poznámka 7. Tyto rovnice ukazují, jak lze vypočítat inverzní matici k matici $A - \boldsymbol{u}\boldsymbol{v}^T$, resp. $A - UV^T$ bez explicitního výpočtu této inverzní matice, známe-li matici A^{-1}.

Příklad 4.7. Je dána matice $A \in \mathcal{M}_n$ a matice k ní inverzní $A^{-1} \in \mathcal{M}_n$:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix} \quad \text{a} \quad A^{-1} = \begin{pmatrix} -3 & -1 & 1 \\ 14 & 2 & -3 \\ -10 & -1 & 2 \end{pmatrix}.$$

Nechť $\boldsymbol{u} = \boldsymbol{v} = (1,0,0)^T$. Odtud

$$A - \boldsymbol{u}\boldsymbol{v}^T = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix}.$$

Nyní

$$\alpha = \frac{1}{1 - \boldsymbol{v}^T A^{-1} \boldsymbol{u}} = \tfrac{1}{4}$$

a

$$(A^{-1} + \alpha A^{-1}\boldsymbol{u}\boldsymbol{v}^T A^{-1}) = (A - \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \begin{pmatrix} -\frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{7}{2} & -\frac{3}{2} & \frac{1}{2} \\ -\frac{5}{2} & \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$

Poznámka 8. Víme, že GEM bez výběru hlavního prvku definuje rozklad matice A ve tvaru

$$A = LR,$$

kde L je dolní trojúhelníková matice s 1 na diagonále, R je horní trojúhelníková matice. Z tohoto vztahu plyne ihned vzorec pro výpočet determinantu matice A, neboť

$$\det A = \det L \det R = \det U = \prod_{i=1}^{n} a_{ii}^{(i)}.$$

Výpočet pomocí GEM s částečným výběrem hlavního prvku vede na rozklad

$$PA = LR,$$

kde $\det P = (-1)^r$, r je počet výměn řádků během výpočtu. Odtud

$$\det PA = (-1)^r \det A = \prod_{i=1}^{n} a_{ii}^{(i)},$$

neboli

$$\det A = (-1)^r \prod_{i=1}^{n} a_{ii}^{(i)}.$$

§ 4.5. Metody založené na minimalizaci kvadratické formy

V tomto odstavci se budeme zabývat metodami, které jsou založeny na minimalizaci kvadratické funkce, jejímž jediným minimem je řešení rovnice $A\boldsymbol{x} = \boldsymbol{b}$. Budeme předpokládat, že matice A je symetrická a pozitivně definitní.

Věta 4.9. *Jestliže A je pozitivně definitní matice, pak řešení systému $A\boldsymbol{x} = \boldsymbol{b}$ je ekvivalentní minimalizaci kvadratické funkce*

$$Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}. \tag{4.22}$$

Tato kvadratická funkce má jediné minimum, kterého nabývá v řešení systému $A\boldsymbol{x} = \boldsymbol{b}$, tj. pro $\boldsymbol{x}^* = A^{-1}\boldsymbol{b}$.

Důkaz. Jednoznačnost minima plyne z pozitivní definitnosti matice A. Nechť $\boldsymbol{x}^* = A^{-1}\boldsymbol{b}$. Uvažujme rozdíl $Q(\boldsymbol{x}^* + \Delta\boldsymbol{x}) - Q(\boldsymbol{x}^*)$, kde $\Delta\boldsymbol{x} \neq \boldsymbol{o}$. Pro tento rozdíl platí

$$\begin{aligned} Q(\boldsymbol{x}^* + \Delta\boldsymbol{x}) - Q(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{x}^* + \Delta\boldsymbol{x})^T A(\boldsymbol{x}^* + \Delta\boldsymbol{x}) - (\boldsymbol{x}^* + \Delta\boldsymbol{x})^T\boldsymbol{b} - \\ - \frac{1}{2}\boldsymbol{x}^{*T} A\boldsymbol{x}^* + \boldsymbol{x}^{*T}\boldsymbol{b} = \frac{1}{2}\Delta\boldsymbol{x}^T A\Delta\boldsymbol{x}. \end{aligned}$$

Matice A je pozitivně definitní a tudíž $\Delta\boldsymbol{x}^T A\Delta\boldsymbol{x} > 0$ pro $\Delta\boldsymbol{x} \neq \boldsymbol{o}$. To znamená, že $Q(\boldsymbol{x}^* + \Delta\boldsymbol{x}) - Q(\boldsymbol{x}^*) > 0$, tj. $Q(\boldsymbol{x}^* + \Delta\boldsymbol{x}) > Q(\boldsymbol{x}^*)$, a tedy $\boldsymbol{x}^*$ realizuje minimum kvadratické funkce Q.

Nechť nyní $\hat{\boldsymbol{x}}$ realizuje minimum kvadratické funkce Q. Nechť dále $\boldsymbol{v} \in \mathbb{R}^n$ je libovolný vektor. Uvažujme vektory tvaru $\boldsymbol{z} = \hat{\boldsymbol{x}} + t\boldsymbol{v}$, t je reálné číslo. Tyto vektory leží na přímce vycházející z $\hat{\boldsymbol{x}}$. Vypočtěme hodnotu funkce Q pro $\boldsymbol{z}$. Je

$$\begin{aligned} Q(\hat{\boldsymbol{x}} + t\boldsymbol{v}) &= \frac{1}{2}(\hat{\boldsymbol{x}} + t\boldsymbol{v})^T A(\hat{\boldsymbol{x}} + t\boldsymbol{v}) - (\hat{\boldsymbol{x}} + t\boldsymbol{v})^T\boldsymbol{b} = \\ &= \frac{1}{2}\hat{\boldsymbol{x}}^T A\hat{\boldsymbol{x}} + \frac{1}{2}t\boldsymbol{v}^T A\hat{\boldsymbol{x}} + \frac{1}{2}t\hat{\boldsymbol{x}}^T A\boldsymbol{v} \; + \\ &\quad + \frac{1}{2}t^2\boldsymbol{v}^T A\boldsymbol{v} \; - \hat{\boldsymbol{x}}\boldsymbol{b}^T - t\boldsymbol{v}^T\boldsymbol{b}. \end{aligned}$$

Jelikož A je pozitivně definitní, je $\hat{\boldsymbol{x}}^T A\boldsymbol{v} = \boldsymbol{v}^T A\hat{\boldsymbol{x}}$. Odtud

$$Q(\hat{\boldsymbol{x}} + t\boldsymbol{v}) = \frac{1}{2}\hat{\boldsymbol{x}}^T A\hat{\boldsymbol{x}} + t\boldsymbol{v}^T A\hat{\boldsymbol{x}} + \frac{1}{2}t^2\boldsymbol{v}^T A\boldsymbol{v} \; - \hat{\boldsymbol{x}}^T\boldsymbol{b} - t\boldsymbol{v}^T\boldsymbol{b}.$$

Funkce Q má minimum v bodě $\hat{\boldsymbol{x}}$, to znamená, že

$$\left.\frac{\mathrm{d}Q(\hat{\boldsymbol{x}} + t\boldsymbol{v})}{\mathrm{d}t}\right|_{t=0} = 0,$$

kde

$$\frac{\mathrm{d}Q(\hat{\boldsymbol{x}} + t\boldsymbol{v})}{\mathrm{d}t} = \boldsymbol{v}^T A\hat{\boldsymbol{x}} + t\boldsymbol{v}^T A\boldsymbol{v} - \boldsymbol{v}^T\boldsymbol{b}.$$

A odtud

$$\begin{aligned} \boldsymbol{v}^T A\hat{\boldsymbol{x}} - \boldsymbol{v}^T\boldsymbol{b} &= 0 \\ \boldsymbol{v}^T(A\hat{\boldsymbol{x}} - \boldsymbol{b}) &= 0. \end{aligned}$$

Vektor $\boldsymbol{v}$ je libovolný vektor z $\mathbb{R}^n$. Poslední vztah znamená, že vektor $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ je ortogonální ke všem vektorům $\boldsymbol{v} \in \mathbb{R}^n$ a odtud plyne, že $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \boldsymbol{o}$. Vektor $\hat{\boldsymbol{x}}$ tedy je řešením systému $A\boldsymbol{x} = \boldsymbol{b}$, a protože A je pozitivně definitní, je $\hat{\boldsymbol{x}} = \boldsymbol{x}^*$. □

Důsledek. $\min Q(\boldsymbol{x}) = Q(\boldsymbol{x}^*) = -\frac{1}{2}\boldsymbol{b}^T A^{-1}\boldsymbol{b}$.

Důkaz. $Q(\boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{x}^{*T} A\ \boldsymbol{x}^* - \boldsymbol{x}^{*T}\boldsymbol{b}$, $\boldsymbol{x}^*$ je přesným řešením systému $A\boldsymbol{x} = \boldsymbol{b}$, a tedy $A\boldsymbol{x}^* = \boldsymbol{b}$, $\boldsymbol{x}^{*T} = \boldsymbol{b}^T A^{-1}$. Dosazením do předchozího vztahu ihned plyne tvrzení. □

Podívejme se nyní na geometrickou interpretaci hledání minima kvadratického funkcionálu Q (viz [12], [19]). Nechť $\boldsymbol{x}^*$ je řešení systému $A\boldsymbol{x} = \boldsymbol{b}$ a nechť $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{s}$. Pak

$$Q(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^{*T}\boldsymbol{b} = \frac{1}{2}(\boldsymbol{x}^* + \boldsymbol{s})^T A\ (\boldsymbol{x}^* + \boldsymbol{s}) - (\boldsymbol{x}^* + \boldsymbol{s})^T\boldsymbol{b} + \frac{1}{2}\boldsymbol{x}^{*T}\boldsymbol{b} = \frac{1}{2}\boldsymbol{s}^T A\boldsymbol{s}\,.$$

Plocha

$$S = \{\boldsymbol{s} = (s_1, \ldots, s_n)^T | \frac{1}{2}\boldsymbol{s}^T A\boldsymbol{s} = konst.\}$$

je hyperelipsoid v proměnných $s_1,\ldots,s_n$ se středem v $\boldsymbol{s} = \boldsymbol{o}$, tj. v $\boldsymbol{x}^*$. Tedy i rovnice $Q(\boldsymbol{x}) = konst.$ představuje hyperelipsoid. Protože A je pozitivně definitní existuje ortogonální matice P taková, že matice

$$P^T AP = D$$

je diagonální matice s kladnými vlastními čísly λ_i matice A na diagonále. Provedeme-li transformaci proměnných

$$\boldsymbol{z} = P^T\boldsymbol{s}, \quad \boldsymbol{z} = (z_1, \ldots, z_n)^T,$$

dostaneme

$$\boldsymbol{s}^T A\ \boldsymbol{s} = \boldsymbol{z}^T D\boldsymbol{z} = \sum_{i=1}^{n} \lambda_i z_i^2\,.$$

Odtud plyne, že hyperelipsoid má své osy ve směrech z_i a délky těchto os jsou přímo úměrné $\frac{1}{\sqrt{\lambda_i}}$, $i = 1, \ldots, n$.

Tyto geometrické úvahy budeme nyní ilustrovat na jednoduchém příkladě.

Příklad 4.8. Uvažujme systém

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

jehož přesné řešení je $\boldsymbol{x}^* = (1,1)^T$.

$$Q(\boldsymbol{x}) = \frac{1}{2}(2x_1^2 + 2x_2^2 - 2x_1x_2) - (x_1 + x_2)$$

a

$$Q(\boldsymbol{x}^*) = -1\,.$$

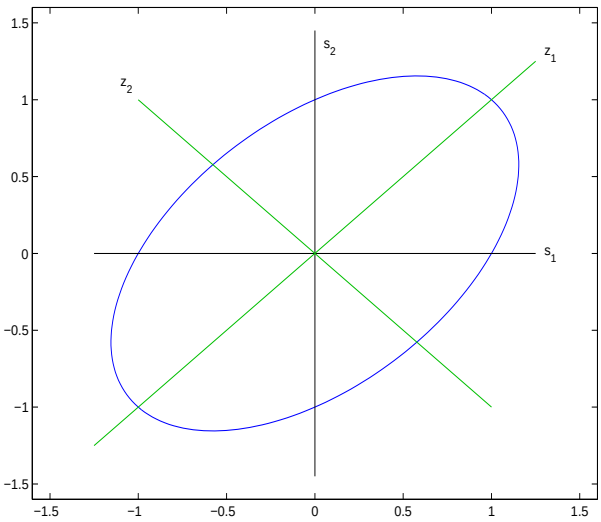
Rovnice elipsy je dána vztahem $\frac{1}{2}\boldsymbol{s}^T A\boldsymbol{s} = c$, c = konstanta

$$\frac{1}{2}(2s_1^2 + 2s_2^2 - 2s_1 s_2) = c.$$

Obrázek 4.1 ukazuje tvar elipsy pro $c = 1$.

Vlastní čísla matice A jsou $\lambda_1 = 3$, $\lambda_2 = 1$. Tedy rovnice příslušné elipsy je

$$\frac{z_1^2}{\sqrt{3}} + \frac{z_2^2}{1} = \tilde{c}\,.$$



Obr. 4.1: Elipsa $\frac{1}{2}(2s_1^2 + 2s_2^2 - 2s_1 s_2) = 1$.

Vlastní vektor příslušný vlastnímu číslu $\lambda_1 = 3$ je tvaru $\boldsymbol{x}^{(1)} = \frac{1}{\sqrt{2}}(-1, 1)^T$ a vektor příslušný vlastnímu číslu $\lambda_2 = 1$ je $\boldsymbol{x}^{(2)} = \frac{1}{\sqrt{2}}(1, 1)^T$, takže

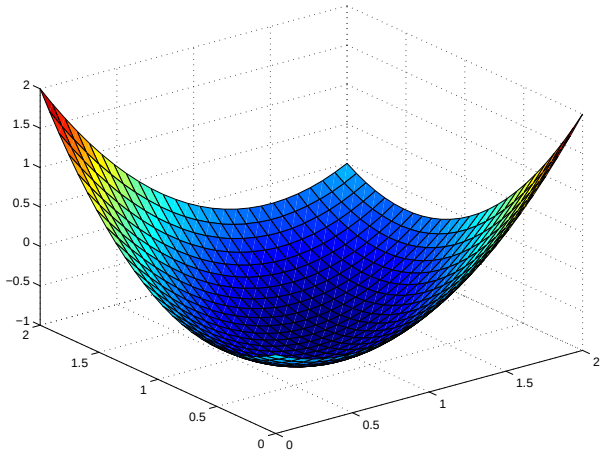
$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Na obrázku 4.2 vidíme průběh kvadratického funkcionálu $Q(\boldsymbol{x})$.

Obecné schema minimalizace funkce Q bude takové, že vybereme počáteční aproximaci $\boldsymbol{x}^1$ a pak určíme $\boldsymbol{x}^2$ tak, že zvolíme nějaký směr $\boldsymbol{v}^1$ a vzdálenost t_1 ve směru $\boldsymbol{v}^1$. Obecně pak

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + t_k \boldsymbol{v}^k, \quad i = 1, 2, \ldots$$

Vektory $\boldsymbol{v}^k$ se nazývají *směrové vektory.*



Obr. 4.2: Průběh kvadratického funkcionálu $Q(\boldsymbol{x})$, $Q(\boldsymbol{x}^*) = -1$

Popišme nyní jednu z metod tohoto typu, a to metodu *největšího spádu*.

Nechť $\boldsymbol{x}^1 \in \mathbb{R}^n$ je počáteční aproximace přesného řešení $\boldsymbol{x}^*$ a položme $\boldsymbol{r}^1 = \boldsymbol{b} - A\boldsymbol{x}^1$; vektor $\boldsymbol{r}^1$ se nazývá *reziduový vektor*. Našim cílem je najít takový vektor $\boldsymbol{v}^1 \in \mathbb{R}^n$, $\|\boldsymbol{v}^1\| = \|A\boldsymbol{x}^1 - \boldsymbol{b}\|$, pro který

$$\frac{\mathrm{d}}{\mathrm{d}t} Q(\boldsymbol{x}^1 + t\boldsymbol{v}^1)\Big|_{t=0} = max, \quad t \in \mathbb{R}\,.$$

Geometricky lze tuto metodu vysvětlit takto:
$\boldsymbol{x} = \boldsymbol{x}^1 + t\boldsymbol{v}^1$, $t \in \mathbb{R}$, je přímka procházející body $\boldsymbol{x}^1$ a $\boldsymbol{v}^1$. Kvadratická funkce Q je plocha v $\mathbb{R}^n$ a bodům ležícím na uvedené přímce odpovídá křivka na této ploše. Nyní hedáme takový směr, tj. takový vektor $\boldsymbol{v}^1$, ve kterém má plocha největší spád.

Vypočtěme hodnotu funkce Q pro body na přímce $\boldsymbol{x} = \boldsymbol{x}^1 + t\boldsymbol{v}$:

$$Q(\boldsymbol{x}^1 + t\boldsymbol{v}) = \frac{1}{2}{\boldsymbol{x}^1}^T A\boldsymbol{x}^1 + t{\boldsymbol{v}^1}^T(A\boldsymbol{x}^1 - \boldsymbol{b}) + \frac{1}{2}t^2{\boldsymbol{v}^1}^T A\boldsymbol{v}^1 - {\boldsymbol{x}^1}^T\boldsymbol{b}\,.$$

Dále

$$\frac{\mathrm{d}Q(\boldsymbol{x}^1 + t\boldsymbol{v}^1)}{\mathrm{d}t} = {\boldsymbol{v}^1}^T(A\boldsymbol{x}^1 - \boldsymbol{b}) + t{\boldsymbol{v}^1}^T A\boldsymbol{v}^1 \tag{4.23}$$

a pro $t = 0$ dostaneme

$$\frac{\mathrm{d}}{\mathrm{d}t} Q(\boldsymbol{x}^1 + t\boldsymbol{v}^1)\Big|_{t=0} = {\boldsymbol{v}^1}^T(A\boldsymbol{x}^1 - \boldsymbol{b})\,.$$

Při předepsané normě bude skalární součin maximální v případě, že $\boldsymbol{v}^1 = A\boldsymbol{x}^1 - \boldsymbol{b}$. To znamená, že ve směru $\boldsymbol{v}^1 = -\boldsymbol{r}^1$ má plocha Q největší spád.

Nyní musíme v tomto směru najít takový bod $\boldsymbol{x}^2$, pro který Q nabývá minimální hodnoty. Budeme tedy hledat minimum kvadratické funkce $Q(\boldsymbol{x}^1 - t\boldsymbol{r}^1)$ jedné proměnné t. Ze vztahu $\mathrm{d}Q(\boldsymbol{x}^1 - t\boldsymbol{r}^1)/\mathrm{d}t = 0$ plyne

$$\boldsymbol{r}^{1T}\boldsymbol{r}^1 + t\boldsymbol{r}^{1T}A\ \boldsymbol{r}^1 = 0$$

tj.

$$t_1 = -\frac{\boldsymbol{r}^{1T}\boldsymbol{r}^1}{\boldsymbol{r}^{1T}A\boldsymbol{r}^1}\,.$$

Funkce $Q(\boldsymbol{x}^1 - t\boldsymbol{r}^1)$ je konvexní vzhledem k proměnné t, neboť

$$\frac{d^2Q(\boldsymbol{x}^1 - t\boldsymbol{r}^1)}{dt^2} = \boldsymbol{v}^{1T}A\boldsymbol{v}^1 > 0.$$

To znamená, že v bodě t_1 se realizuje jediné minimum. Matice A je pozitivně definitní a $\boldsymbol{r}^1 \neq \boldsymbol{o}$, a tedy $\boldsymbol{r}^{1T}A\boldsymbol{r}^1 > 0$. Další aproximace je tvaru

$$\boldsymbol{x}^2 = \boldsymbol{x}^1 + \frac{\boldsymbol{r}^{1T}\boldsymbol{r}^1}{\boldsymbol{r}^{1T}A\boldsymbol{r}^1}\boldsymbol{r}^1\,.$$

Uvedená metoda se nazývá metoda *největšího spádu* a její algoritmus má obecně tvar:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \frac{\boldsymbol{r}^{kT}\boldsymbol{r}^k}{\boldsymbol{r}^{kT}A\boldsymbol{r}^k}\boldsymbol{r}^k\,. \tag{4.24}$$

Nyní odvodíme důležité vztahy mezi reziduálními vektory $\boldsymbol{r}^k$. Z (4.24) plyne

$$A\boldsymbol{x}^{k+1} = A\boldsymbol{x}^k + \frac{\boldsymbol{r}^{kT}\boldsymbol{r}^k}{\boldsymbol{r}^{kT}A\boldsymbol{r}^k}A\boldsymbol{r}^k, \quad \boldsymbol{r}^k = \boldsymbol{b} - A\boldsymbol{x}^k$$

a

$$A\boldsymbol{x}^{k+1} - \boldsymbol{b} = A\boldsymbol{x}^k - \boldsymbol{b} + \frac{\boldsymbol{r}^{kT}\boldsymbol{r}^k}{\boldsymbol{r}^{kT}A\boldsymbol{r}^k}A\boldsymbol{r}^k, \quad \boldsymbol{r}^k = \boldsymbol{b} - A\boldsymbol{x}^k\,.$$

Odtud

$$\boldsymbol{r}^{k+1} = \boldsymbol{r}^k - \frac{\boldsymbol{r}^{kT}\boldsymbol{r}^k}{\boldsymbol{r}^{kT}A\boldsymbol{r}^k}A\boldsymbol{r}^k, \quad \boldsymbol{r}^k = \boldsymbol{b} - A\boldsymbol{x}^k$$

neboli

$$\boldsymbol{r}^{k+1} = \boldsymbol{r}^k + t_kA\boldsymbol{r}^k\,.$$

Dále pro skalární součin vektorů $\boldsymbol{r}^{k+1}$ a $\boldsymbol{r}^k$ platí

$$\boldsymbol{r}^{k+1T}\boldsymbol{r}^k = (\boldsymbol{r}^k + t_kA\boldsymbol{r}^k)^T\boldsymbol{r}^k = \boldsymbol{r}^{kT}\boldsymbol{r}^k + t_k\boldsymbol{r}^{kT}A\boldsymbol{r}^k =$$

$$= \boldsymbol{r}^{kT}\boldsymbol{r}^k - \frac{\boldsymbol{r}^{kT}\boldsymbol{r}^k}{\boldsymbol{r}^{kT}A\boldsymbol{r}^k}\boldsymbol{r}^{kT}A\boldsymbol{r}^k = 0\,,$$

což znamená, že vektory $\boldsymbol{r}^{k+1}$ a $\boldsymbol{r}^k$ jsou ortogonální.

Ukážeme, že pro posloupnost generovanou metodou největšího spádu platí:

$$Q(\boldsymbol{x}^{k+1}0 < Q(\boldsymbol{x}^k), \quad k = 1, 2 \dots$$

Počítejme rozdíl

$$\begin{aligned}
& Q(\boldsymbol{x}^{k+1}) - Q(\boldsymbol{x}^k) = Q\left(\boldsymbol{x}^k + \frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\boldsymbol{r}^k\right) - Q(\boldsymbol{x}^k) = \\
= \quad & \frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\boldsymbol{r}^{k^T} A\boldsymbol{x}^k - \frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\boldsymbol{r}^{k^T}\boldsymbol{b} + \frac{1}{2}\left(\frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\right)^2 \boldsymbol{r}^{k^T} A\boldsymbol{r}^k = \\
= \quad & \frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\left(\boldsymbol{r}^{k^T}(A\boldsymbol{x}^k - b) + \frac{1}{2}\frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\boldsymbol{r}^{k^T} A\boldsymbol{r}^k\right) = \\
= \quad & \frac{\boldsymbol{r}^{k^T}\boldsymbol{r}^k}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k}\left(-\boldsymbol{r}^{k^T}\boldsymbol{r}^k + \frac{1}{2}\boldsymbol{r}^{k^T}\boldsymbol{r}^k\right) = -\frac{1}{2}\frac{(\boldsymbol{r}^{k^T}\boldsymbol{r}^k)^2}{\boldsymbol{r}^{k^T} A\boldsymbol{r}^k} < 0 \;\Rightarrow \\
\Rightarrow \quad & Q(\boldsymbol{x}^{k+1}) < Q(\boldsymbol{x}^k)\,.
\end{aligned}$$

Odtud také plyne konvergence metody největšího spádu (viz [19]).

Příklad 4.9. Řešme metodou největšího spádu systém

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

jehož přesné řešení je $\boldsymbol{x}^* = (1,1)^T$. Zvolme počáteční aproximaci $\boldsymbol{x}^1 = (0,1)^T$, je $Q(\boldsymbol{x}^1) = 0$. Pak

$$\boldsymbol{r}^1 = \boldsymbol{b} - A\boldsymbol{x}^1 = (2,-1)^T \text{ a } \boldsymbol{r}^{1^T}\boldsymbol{r}^1 = 5,\ \boldsymbol{r}^{1^T} A\boldsymbol{r}^1 = 14\,.$$

Další aproximace $\boldsymbol{x}^2$ je tvaru

$$\boldsymbol{x}^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{5}{14}\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 5/7 \\ 9/14 \end{pmatrix}; \quad Q(\boldsymbol{x}^2) = -\frac{30}{49}\,.$$

Dále

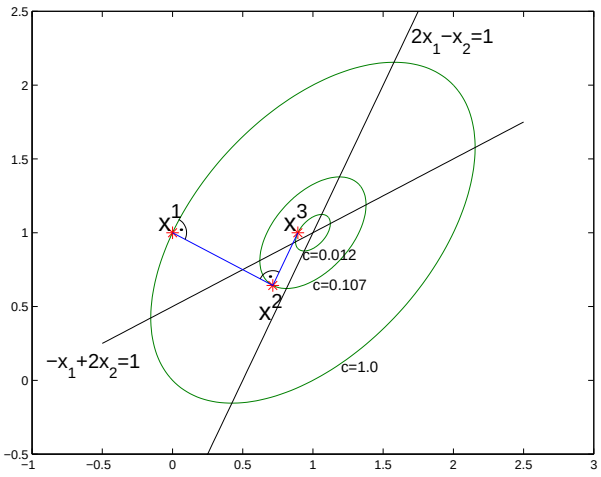
$$\boldsymbol{r}^2 = \boldsymbol{b} - A\boldsymbol{x}^2 = (3/14, 3/7)^T,\ \boldsymbol{r}^{2^T}\boldsymbol{r}^2 = \frac{45}{196},\ \boldsymbol{r}^{2^T} A\ \boldsymbol{r}^2 = \frac{54}{196}\,,$$

odtud

$$\boldsymbol{x}^3 = \begin{pmatrix} 5/7 \\ 9/14 \end{pmatrix} + \frac{45}{54}\begin{pmatrix} 3/14 \\ 3/7 \end{pmatrix} = \begin{pmatrix} \frac{25}{28} \\ 1 \end{pmatrix} \doteq \begin{pmatrix} 0,89826 \\ 1 \end{pmatrix},$$

$Q(\boldsymbol{x}^3) \doteq -0,988518$. První tři iterace ukazuje obrázek 4.3 spolu s elipsami o rovnici $\frac{1}{2}(2s_1^2 + 2s_2^2 - 2s_1 s_2) = c$, na nichž jednotlivé iterace leží.

Další metodu patřící do této skupiny je *metodu sdružených gradientů*. Tato metoda byla navržena v roce 1952 Hestenesem a Stiefelem a nyní se velmi často



Obr. 4.3: Metoda největšího spádu

užívá pro řešení velkých a řídkých systémů s pozitivně definitními maticemi. Z hlediska teoretického je tato metoda přímá, ale při praktické realizaci je to metoda iterační. Podrobný pos této metody lze nalézt např. v [19].

Podívejme se nyní na otázky stability algoritmů a vlivu zaokrouhlovacích chyb při řešení systému $A\boldsymbol{x} = \boldsymbol{b}$.

§ 4.6. Stabilita, podmíněnost

V úvodní kapitole jsme uvedli definici stability. Nyní podrobně probereme otázky stability při řešení systémů lineárních rovnic.

Definice 4.6. Algoritmus pro řešení $A\boldsymbol{x} = \boldsymbol{b}$ se nazývá *stabilní*, jestliže vypočtené řešení $\hat{\boldsymbol{x}}$ je takové, že

$$(A + \mathcal{E})\hat{\boldsymbol{x}} = \boldsymbol{b} + \delta\boldsymbol{b},$$

kde $\mathcal{E}$ a $\delta\boldsymbol{b}$ jsou malé; $\mathcal{E}$ se nazývá *chybová matice*.

Poznámka 9. „Malost“ matice nebo vektoru lze „měřit“ buď na základě jejich prvků nebo výpočtem normy.

Podívejme se nyní na GEM bez výběru pivota. Lze ukázat ([6]), že vypočtené řešení $\hat{\boldsymbol{x}}$ vyhovuje systému

$$(A + \mathcal{E})\hat{\boldsymbol{x}} = \boldsymbol{b},$$

přičemž

$$\|\mathcal{E}\|_\infty \leq cn^3\varrho\|A\|_\infty\mu + O(\mu^2),$$

kde $A^{(k)} = (a_{ij}^{(k)})$, $A^{(1)} = A$, jsou redukované matice v eliminačním procesu, μ je strojová přesnost ($\mu = 10^{-6}$ pro jednoduchou přesnost, $\mu = 10^{-16}$ pro dvojnásobnou přesnost), c je konstanta a ϱ se nazývá růstový faktor a je dán vztahem

$$\varrho = \frac{\max\limits_{k} \max\limits_{i,j} \left|a_{ij}^{(k)}\right|}{\max\limits_{i,j} \left|a_{ij}^{(1)}\right|}.$$

Jestliže $\alpha_1 = \max\limits_{i,j} |a_{i,j}^{(1)}|$, $\alpha_k = \max\limits_{i,j} |a_{i,j}^{(k)}|$, pak ϱ lze vyjádřit takto:

$$\varrho = \frac{\max(\alpha_1, \alpha_2, \ldots, \alpha_n)}{\alpha_1}.$$

Pro libovolnou matici mohou prvky matic $A^{(k)}$ růst libovolně a tedy i faktor ϱ může být velký (podrobněji viz [6]). Ilustrujme tento fakt na příkladu:

Příklad 4.10.

$$A = A^{(1)} = \begin{pmatrix} 10^{-10} & 1 \\ 1 & 2 \end{pmatrix} \quad \rightarrow \quad A^{(2)} = \begin{pmatrix} 10^{-10} & 1 \\ 0 & -10^{10} \end{pmatrix}$$

Růstový faktor

$$\varrho = \frac{\max(\alpha_1, \alpha_2)}{\alpha_1} = \frac{\max\left(2, 10^{10}\right)}{2} = \frac{10^{10}}{2}.$$

Řešíme-li lineární systém s touto maticí, nemůžeme očekávat malou chybovou matici $\mathcal{E}$.

Řešme např. systém

$$\begin{aligned} 10^{-10}x_1 + \ \ x_2 &= 1 \\ x_1 + 2x_2 &= 3 \end{aligned}$$

Užitím $A^{(2)}$ vypočteme $x_1 = 0$, $x_2 = 1$, zatímco přesné řešení je $x_1 = 1$, $x_2 = 1$ (při zaokrouhlování na 9 cifer).

Při GEM je vektor $\boldsymbol{b} = \boldsymbol{b}^{(1)}$ modifikován na vektor $\boldsymbol{b}^{(2)}$: $\boldsymbol{b}^{(2)} = (1, 3 - 10^{10})^T = (1, -10^{10})^T$. Tento fakt ukazuje, že *GEM bez výběru pivota* je obecně nestabilní procedura. Ale na druhé straně může být tato procedura stabilní pro některé speciální typy matice.

Následující tabulky ukazují srovnání různých metod z hlediska stability ([6]).

I. Matice systému je libovolná		
Metoda	**Počet nás. operací**	**Stabilita**
GEM bez výběru pivota	$\frac{n^3}{3}$	nestabilní
GEM s částečným výběrem pivota	$\frac{n^3}{3} + O(n^2)$ porovnání	stabilní
GEM s úplným výběrem pivota	$\frac{n^3}{3} + O(n^3)$ porovnání	stabilní

II. Speciální typy matic			
Matice	**Metoda**	**Počet nás. operací**	**Stabilita**
Symetrická	GEM bez výběru pivota	$\frac{n^3}{3}$	stabilní
Symetrická	Choleského	$\frac{n^3}{6} + n$ druhých odmocnin	stabilní
Diagonálně dominantní	GEM bez výběru pivota	$\frac{n^3}{3}$	stabilní
Třídiagonální	Croutova	$O(n)$	stabilní

Z předchozích úvah bychom neměli nabýt dojmu, že stabilita algoritmu zaručuje, že vypočtené řešení bude přesné. Vlastnost, která se nazývá *podmíněnost*, ale také přispívá k přesnosti nebo nepřesnosti vypočteného výsledku.

Podmíněnost problému je vlastnost problému samotného. Jak jsme již uvedli v úvodu, podmíněnost se týká toho, jak se řešení změní, jestliže se změní vstupní data. Tento problém nastává při praktických aplikacích, kdy vstupní data získaná měřením nebo pozorováním jsou zatížena chybami. Ve skutečnosti tedy musíme řešit problém, který není zadán původními daty, ale daty s „poruchami“. Otázkou tedy je, jaký vliv mají tyto poruchy na řešení. Ilustrujme tento fakt na příkladě.

Příklad 4.11. Předpokládejme, že v nějakém podniku jsou dvě oddělení. V prvním oddělení pracuje 101 žena a 10 mužů, v druhém oddělení 10 žen a 1 muž. Nechť první oddělení dostane za časovou jednotku 111 Kč, druhé oddělení 11 Kč. Ptáme se, jaká je mzda ženy a muže za časovou jednotku?

Označíme-li x_1 mzdu ženy a x_2 mzdu muže, vede úloha na systém rovnic

$$\begin{aligned} 101x_1 + 10x_2 &= 111 \\ 10x_1 + x_2 &= 11 \end{aligned}$$

Je zřejmé, že řešení tohoto systému je: $x_1 = 1$, $x_2 = 1$. Ale vedoucí se rozhodl druhému oddělení přidat, aby posílil „pozici“ jediného muže a zvýšil částku na 11,10 Kč. Systém rovnic je tvaru

$$\begin{aligned} 101x_1 + 10x_2 &= 111 \\ 10x_1 + \quad x_2 &= \ \ 11{,}1 \end{aligned}$$

Ovšem řešení tohoto systému je $\tilde{x}_1 = 0$, $\tilde{x}_2 = 11{,}1$. To znamená, že malá změna na vstupu (pravé strany) má za následek velkou změnu na výstupu. Spočítejme číslo podmíněnosti podle vztahu uvedeného v úvodu.

$$C_p = \frac{\|\text{relativní chyba na výstupu}\|}{\|\text{relativní chyba na vstupu}\|}.$$

K vyjádření tohoto čísla použijeme norem vektorů. Nechť

$$\boldsymbol{x}^* = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \|\boldsymbol{x}^*\|_1 = |x_1| + |x_2| = 2; \qquad \tilde{\boldsymbol{x}} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \Rightarrow \|\tilde{\boldsymbol{x}}\|_1 = 11{,}1$$

$$\boldsymbol{b} = \begin{pmatrix} 111 \\ 11 \end{pmatrix} \Rightarrow \|\boldsymbol{b}\|_1 = 122, \qquad \tilde{\boldsymbol{b}} = \begin{pmatrix} 111 \\ 11{,}1 \end{pmatrix} \Rightarrow \|\tilde{\boldsymbol{b}}\|_1 = 122{,}1$$

Nyní

$$C_p = \frac{\dfrac{\|\boldsymbol{x}^* - \tilde{\boldsymbol{x}}\|_1}{\|\boldsymbol{x}^*\|_1}}{\dfrac{\|\boldsymbol{b} - \tilde{\boldsymbol{b}}\|_1}{\|\boldsymbol{b}\|_1}} = \frac{\dfrac{11{,}1}{2}}{\dfrac{0{,}1}{122}} \approx 6770.$$

Jedná se o špatně podmíněnou úlohu.

Pro vyšetřování podmíněnosti systémů lineárních rovnic je vhodné definovat číslo podmíněnosti matice.

Definice 4.7. Pro libovolnou přidruženou maticovou normu definujeme *číslo podmíněnosti* matice A vztahem

$$k(A) = \|A\| \, \|A^{-1}\|.$$

Řekneme, že *matice A je dobře podmíněna*, jestliže $k(A) \approx 1$ a *špatně podmíněna*, jestliže $k(A)$ je podstatně větší než 1.

Je jasné, že $k(A) \geq 1$, neboť

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

Vypočtěme číslo podmíněnosti matice A z předchozího příkladu:

$$\begin{aligned} A \quad &= \begin{pmatrix} 101 & 10 \\ 10 & 1 \end{pmatrix} \qquad \Rightarrow \qquad \|A\|_1 \quad = 111 \\ A^{-1} &= \begin{pmatrix} 1 & -10 \\ -10 & 101 \end{pmatrix} \qquad \Rightarrow \qquad \|A^{-1}\|_1 = 111 \end{aligned}$$

Tedy $k(A) = 111^2 = 12321$, což opět znamená, že A je špatně podmíněna.

§ 4.7. Analýza chyb

Nechť nyní matice A je dána s poruchou δA a vektor $\boldsymbol{b}$ s poruchou $\delta\boldsymbol{b}$. Tedy místo systému $A\boldsymbol{x} = \boldsymbol{b}$ řešíme systém $(A + \delta A)(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}$.

Věta 4.10. *Nechť A je regulární matice a nechť pro nějakou přidruženou maticovou normu platí:*

$$\|\delta A\| < \frac{1}{\|A^{-1}\|}.$$

Řešení $\tilde{\boldsymbol{x}} = \boldsymbol{x}^* + \delta\boldsymbol{x}^*$ systému $(A + \delta A)(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}$ aproximuje řešení $\boldsymbol{x}^*$ systému $A\boldsymbol{x} = \boldsymbol{b}$ s chybou

$$\frac{\|\delta\boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq \frac{k(A)}{1 - k(A)\dfrac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\delta A\|}{\|A\|} \right). \tag{4.25}$$

Důkaz. Z předpokladu $\|\delta A\| < 1/\|A^{-1}\|$ plyne

$$\|A^{-1}\delta A\| \leq \|A^{-1}\| \, \|\delta A\| < 1\,,$$

a tedy podle věty 1.5 je matice $(E + A^{-1}\delta A)$ regulární a platí

$$\|(E + A^{-1}\delta A)^{-1}\| \leq \frac{1}{1 - \|A^{-1}\delta A\|} \leq \frac{1}{1 - \|A^{-1}\| \, \|\delta A\|}. \tag{4.26}$$

Upravíme systém

$$(A + \delta A)(\boldsymbol{x}^* + \delta\boldsymbol{x}^*) = \boldsymbol{b} + \delta\boldsymbol{b}$$

následujícím způsobem:

$$\begin{aligned} A^{-1}(A + \delta A)(\boldsymbol{x}^* + \delta\boldsymbol{x}^*) &= A^{-1}(\boldsymbol{b} + \delta\boldsymbol{b}) \\ (E + A^{-1}\delta A)(\boldsymbol{x}^* + \delta\boldsymbol{x}^*) &= A^{-1}(\boldsymbol{b} + \delta\boldsymbol{b}) \\ \boldsymbol{x}^* + \delta\boldsymbol{x}^* &= (E + A^{-1}\delta A)^{-1}A^{-1}(\boldsymbol{b} + \delta\boldsymbol{b}). \end{aligned}$$

Na druhé straně je

$$\boldsymbol{x}^* = (E + A^{-1}\delta A)^{-1}(E + A^{-1}\delta A)\boldsymbol{x}^*$$

a odtud

$$\boldsymbol{x}^* + \delta\boldsymbol{x}^* - \boldsymbol{x}^* = (E + A^{-1}\delta A)^{-1} \left\{ A^{-1}(\boldsymbol{b} + \delta\boldsymbol{b}) - \boldsymbol{x}^* - A^{-1}\delta A\boldsymbol{x}^* \right\},$$

tj.

$$\delta\boldsymbol{x}^* = (E + A^{-1}\delta A)^{-1}A^{-1}(\delta\boldsymbol{b} - \delta A\boldsymbol{x}^*).$$

Přechodem k normě dostaneme

$$\|\delta \boldsymbol{x}^*\| \leq \|(E + A^{-1}\delta A)^{-1}\| \, \|A^{-1}\| \, (\|\delta \boldsymbol{b}\| + \|\delta A\| \, \|\boldsymbol{x}^*\|),$$

neboť přidružená maticová norma je souhlasná s danou vektorovou normou.

V dalších úpravách užijeme vztahu (4.26):

$$\frac{\|\delta \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \, \|\delta A\|} \left(\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{x}^*\|} + \|\delta A\| \right)$$

Jelikož $\|A\boldsymbol{x}^*\| = \|\boldsymbol{b}\|$, je $\|\boldsymbol{b}\| \leq \|A\| \, \|\boldsymbol{x}^*\|$, a tedy $\|\boldsymbol{x}^*\| \geq \|\boldsymbol{b}\|/\|A\|$. Odtud

$$\frac{\|\delta \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq \frac{\|A^{-1}\| \, \|A\|}{1 - \|A^{-1}\| \, \|\delta A\|} \left(\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\delta A\|}{\|A\|} \right).$$

Užijme nyní vyjádření pro číslo podmíněnosti $k(A) = \|A\|.\|A^{-1}\|$ a dostaneme požadovaný odhad (4.25). □

Poznámka 10. Jestliže $\delta \boldsymbol{b} = \boldsymbol{o}$, ukazuje předchozí věta vliv poruchy matice A na relativní chybu řešení. V tomto případě

$$\frac{\|\delta \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq \frac{k(A)}{1 - k(A)\dfrac{\|\delta A\|}{\|A\|}} \frac{\|\delta A\|}{\|A\|}.$$

Jmenovatel zlomku na pravé straně této nerovnosti je menší než 1. Tedy, dokonce i za předpokladu, že $\|\delta A\|/\|A\|$ je malé číslo, chyba řešení může být značná, jestliže $k(A)$ je velké číslo. Stejný závěr platí i v případě existence poruch δA i $\delta \boldsymbol{b}$. *Číslo podmíněnosti $k(A)$ má tedy zásadní význam pokud jde o citlivost řešení vzhledem ke vstupním datům.*

Předpokládejme nyní, že matice A je dána přesně, ale vektor $\boldsymbol{b}$ je dán s poruchami. Tedy místo systému $A\boldsymbol{x} = \boldsymbol{b}$ řešíme systém $A(\boldsymbol{x} + \delta \boldsymbol{x}) = \boldsymbol{b} + \delta \boldsymbol{b}$.

Věta 4.11. *Nechť A je regulární matice a vektor $\boldsymbol{b} \neq \boldsymbol{o}$. Jestliže $\delta \boldsymbol{b}$ resp. $\delta \boldsymbol{x}$ jsou poruchy vektoru $\boldsymbol{b}$ resp. $\boldsymbol{x}$, pak*

$$\frac{\|\delta \boldsymbol{b}\|}{k(A)\|\boldsymbol{b}\|} \leq \frac{\|\delta \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq k(A) \frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}.$$

Důkaz pravé nerovnosti je důsledkem věty 4.10 a důkaz nerovnosti vlevo lze nalézt v [6]. Zde uvedeme interpretaci věty. Tato věta říká, že *relativní chyba řešení může být tak velká jako číslo podmíněnosti matice A násobené relativní chybou vektoru $\boldsymbol{b}$*. Jestliže číslo podmíněnosti není příliš velké, potom malé změny vektoru $\boldsymbol{b}$ mají za následek malé změny řešení. Vraťme se k předchozímu příkladu. Zde je

$$\frac{\|\delta \boldsymbol{x}^*\|_1}{\|\boldsymbol{x}^*\|_1} \leq 111^2 . \frac{0{,}1}{122} \approx 10{,}1, \qquad \boldsymbol{x}^* = (x_1, x_2)^T.$$

Skutečná relativní chyba je $11{,}2/2 = 5{,}6$.

Je třeba poznamenat, že horní hranice chyby je podstatně vyšší. Tento fakt vyplývá z celkové koncepce odhadu chyb v numerické matematice, kdy vždy uvažujeme horní odhad chyb, tj. nejhorší možný případ.

Příklad 4.12. Velmi známým příkladem špatně podmíněné matice je Hilbertova matice:

$$A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{pmatrix}.$$

Pro $n = 10$ vzhledem k normě $\| \, . \, \|_1$ je $k(A) = 3{,}5353.10^{13}$. Je vhodné si povšimnout, že determinant této matice je velmi malý.

Objasníme nyní praktický význam věty 4.10.

Předpokládejme, že výpočet v pohyblivé řádové čárce se zaokrouhlováním na t desetinných míst může zapříčinit relativní chyby dané v normách vztahy:

$$\frac{\|\delta A\|}{\|A\|} \approx 5.10^{-t}, \qquad \frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} \approx 5.10^{-t}.$$

Za předpokladu, že $k(A) \approx 10^{\alpha}$ a $5.10^{\alpha - t} \ll 1$, dostaneme z odhadu ve větě 4.10:

$$\frac{\|\delta \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq 10^{\alpha - t + 1}.$$

Tato skutečnost vede k následujícímu závěru:

Jestliže řešíme systém $A\boldsymbol{x} = \boldsymbol{b}$ v pohyblivé řádové čárce se zaokrouhlováním na t desetinných míst a $k(A) \approx 10^{\alpha}$, pak vypočtené řešení $\tilde{\boldsymbol{x}}$ je správné na $(t - \alpha - 1)$ desetinných míst.

Nechť nyní $\tilde{\boldsymbol{x}}$ je vypočtené řešení (jakoukoliv metodou). Uvažujme reziduový vektor $\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}}$. Zdálo by se logické, že když $\|\boldsymbol{r}\|$ je malé číslo, je $\tilde{\boldsymbol{x}}$ dobrou aproximací přesného řešení. Ale následující příklad ukazuje, že tomu tak být nemusí.

Příklad 4.13. Uvažujme systém

$$\begin{pmatrix} 1 & 2 \\ 1{,}0001 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3{,}0001 \end{pmatrix}.$$

Tento systém má jediné řešení $\boldsymbol{x}^* = (1, 1)^T$. Pro aproximaci $\tilde{\boldsymbol{x}} = (3, 0)^T$ je reziduový vektor tvaru

$$\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}} = \begin{pmatrix} 3 \\ 3{,}0001 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 1{,}0001 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0{,}0002 \end{pmatrix}.$$

Tedy $\|\boldsymbol{r}\|_\infty = 0{,}0002$, ale $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_\infty = 2$.

Z geometrického hlediska lze tuto situaci vysvětlit takto:

Řešení systému je průsečík přímek

$$l_1: \quad x_1 + 2x_2 = 3, \qquad l_2: \quad 1{,}0001x_1 + 2x_2 = 3{,}0001.$$

Bod $\tilde{\boldsymbol{x}} = (3, 0)$ leží na přímce l_1 a přímky l_1 a l_2 jsou téměř rovnoběžné, což implikuje, že bod $(3, 0)$ leží blízko přímky l_2, i když se podstatně liší od průsečíku přímek v bodě $(1, 1)$.

Matematicky lze tento jev objasnit následující větou.

Věta 4.12. *Nechť $\tilde{\boldsymbol{x}}$ je aproximace řešení systému $A\boldsymbol{x} = \boldsymbol{b}$ s regulární maticí A. Pak pro přidruženou maticovou normu platí:*

$$\|\boldsymbol{x}^* - \tilde{\boldsymbol{x}}\| \leq \|\boldsymbol{r}\| \, \|A^{-1}\|, \tag{4.27}$$

$$\frac{\|\boldsymbol{x}^* - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}^*\|} \leq k(A)\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}, \qquad \text{za předpokladu } \boldsymbol{b} \neq \boldsymbol{o}. \tag{4.28}$$

Důkaz. Je $\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}} = A\boldsymbol{x}^* - A\tilde{\boldsymbol{x}} \Rightarrow \boldsymbol{x}^* - \tilde{\boldsymbol{x}} = A^{-1}\boldsymbol{r}$, neboť A je regulární matice.

Z vlastností přidružené normy plyne:

$$\|\boldsymbol{x}^* - \tilde{\boldsymbol{x}}\| = \|A^{-1}\boldsymbol{r}\| \leq \|A^{-1}\| \, \|\boldsymbol{r}\|$$

Dále $\boldsymbol{b} = A\boldsymbol{x}^*$ a tedy $\|\boldsymbol{b}\| \leq \|A\| \, \|\boldsymbol{x}^*\|$, tj. $\|\boldsymbol{x}^*\| \geq \|\boldsymbol{b}\|/\|A\|$. Použitím vztahu (4.27) nyní dostaneme

$$\frac{\|\boldsymbol{x}^* - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}^*\|} \leq \frac{\|A^{-1}\| \, \|A\|}{\|\boldsymbol{b}\|}\|\boldsymbol{r}\| = k(A)\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}.$$

□

Tato věta říká, že *relativní chyba aproximace $\tilde{\boldsymbol{x}}$ závisí nejen na reziduovém vektoru, ale také na čísle podmíněnosti matice A*. Vypočtené řešení tedy bude dostatečně přesné pouze v případě, že součin čísla podmíněnosti a relativního rezidua $\|\boldsymbol{r}\|/\|\boldsymbol{b}\|$ je malé číslo. Vraťme se nyní k předchozímu příkladu a vypočtěme číslo podmíněnosti matice A vzhledem k $\| \, . \, \|_\infty$. Je

$$A^{-1} = \begin{pmatrix} -10000 & 10000 \\ 5000{,}5 & -5000 \end{pmatrix},$$

$\|A\|_\infty = 3{,}0001$, $\|A^{-1}\|_\infty = 20000$, tzn.

$$k(A) = 60002.$$

Číslo podmíněnosti je velmi velké, což znamená, že i v případě, že norma reziduového vektoru je malé číslo, chyba aproximace může být velká.

Poznámka 11. Nechť A je pozitivně definitní matice. Uvažujme spektrální normu $\|A\|_2 = \sqrt{\varrho(A^T A)}$. Protože je A pozitivně definitní, platí

$$\|A\|_2 = \max_{1\le i\le n} \lambda_i$$

a

$$\|A^{-1}\|_2 = (\min_{1\le i\le n} \lambda_i)^{-1}.$$

Číslo podmíněnosti $k(A)$ je v tomto případě tvaru

$$k(A) = \frac{\max \lambda_i}{\min \lambda_i}\,.$$

Vrátíme-li se ke geometrické interpretaci hledání minima kvadratické funkce v odstavci 4.5, je rovnice $\boldsymbol{s}^T A\boldsymbol{s} = c$ rovnicí hyperelipsoidu, jehož kanonický tvar je

$$\sum_{i=1}^{n} \lambda_i z_i^2 = \tilde{c}\,.$$

Je-li tedy matice A dobře podmíněna, $k(A) \approx 1$, jsou hyperelipsoidy blízké kulovým nadplochám a naopak pro $k(A) \gg 1$ jsou hyperelipsoidy protáhlé.

Cvičení ke kapitole 4

1. Řešte systém GEM a) bez výběru hlavního prvku, b) s částečným výběrem hlavního prvku, c) s úplným výběrem hlavního prvku:

$$\begin{aligned} x_1 - x_2 + 2x_3 - x_4 &= -8 \\ 2x_1 - 2x_2 + 3x_3 - 3x_4 &= -20 \\ x_1 + x_2 + x_3 &= -2 \\ x_1 - x_2 + 4x_3 + 3x_4 &= 4 \end{aligned}$$

   (Řešení: $x_1 = -7$, $x_2 = 3$, $x_3 = 2$, $x_4 = 2$.)

2. Užijte Gaussovy eliminační metody s částečným výběrem hlavního prvku pro řešení soustavy

$$\begin{aligned} x_2 + x_3 &= 0 \\ 2x_1 + 2x_2 + 3x_3 &= 1 \\ x_1 + 2x_2 + x_3 &= 5 \end{aligned}\quad .$$

3. Ukažte že matici A nelze rozložit na součin horní a dolní trojúhelníkové matice:

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

   Řešte nyní systémy $A\boldsymbol{x} = \boldsymbol{b}_1$, $A\boldsymbol{x} = \boldsymbol{b}_2$, kde $\boldsymbol{b}_1 = (7, 8, 10, 0)^T$, $\boldsymbol{b}_2 = (7, 5, 10, 0)^T$. Užijte GEM a ukažte, že systém $A\boldsymbol{x} = \boldsymbol{b}_1$ má nekonečně mnoho řešení a systém $A\boldsymbol{x} = \boldsymbol{b}_2$ nemá žádné řešení.

4. Choleského metodou řešte soustavu

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 + 5x_2 + 5x_3 &= 11 \\ x_1 + 5x_2 + 14x_3 &= 20 \end{aligned} .$$

5. Řešte systém $H\boldsymbol{x} = \boldsymbol{b}$ s Hilbertovou maticí H:

   a)

$$H = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\ \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \end{pmatrix} \qquad \boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} .$$

   b) Dále řešte tento systém s maticí $\tilde{H} = H + \delta H$ a porovnejte výsledky

$$H + \delta H = \begin{pmatrix} 1{,}0 & 0{,}5 & 0{,}33333 & 0{,}25 & 0{,}2 \\ 0{,}5 & 0{,}33333 & 0{,}25 & 0{,}2 & 0{,}16667 \\ 0{,}33333 & 0{,}25 & 0{,}2 & 0{,}16667 & 0{,}14286 \\ 0{,}25 & 0{,}2 & 0{,}16667 & 0{,}14286 & 0{,}125 \\ 0{,}2 & 0{,}16667 & 0{,}14286 & 0{,}125 & 0{,}11111 \end{pmatrix}$$

$$\boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} .$$

(a) Řešení s maticí H:
$\boldsymbol{x} = (25, -300, 1050, -1400, 630)^T$.
b) Řešení s maticí $\tilde{H}$:
$\tilde{\boldsymbol{x}} = (28{,}02304; -348{,}5887; 1239{,}781; -1666{,}785; 753{,}5564)^T$.)

6. Přesné řešení systému

$$\begin{aligned} 1{,}133x_1 + 5{,}281x_2 &= 6{,}414 \\ 24{,}14x_1 - 1{,}210x_2 &= 22{,}93 \end{aligned}$$

je $\boldsymbol{x} = (1, 1)^T$. Řešte tento systém se zaokrouhlováním na 4 cifry

   a) GEM bez výběru hlavního prvku,

   b) GEM s částečným výběrem hlavního prvku.

(a) $x_1 = 0{,}9956$, $x_2 = 1{,}001$; b) $x_1 = 1{,}000$, $x_2 = 1{,}000$.)

7. Uvažujme systém

$$\begin{aligned} x_1 + 2x_2 &= 2 \\ 2x_1 + 3x_2 &= 3{,}4\ , \end{aligned}$$

jehož přesné řešení je $\boldsymbol{x} = (0{,}8; 0{,}6)^T$. Vypočtěte reziduový vektor pro aproximaci $\tilde{\boldsymbol{x}} = (1{,}00; 0{,}48)^T$ a vektor skutečné chyby řešení. Výsledky vysvětlete graficky.

8. Vypočtěte matici inverzní k matici

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

(správnost výsledků zkontrolujte výpočtem AA^{-1}).

9. Najděte přímý rozklad $A = LU$ ($l_{ii} = 1$, $i = 1, 2, 3$)

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

(Řešení: $L = \begin{pmatrix} 1 & 0 & 0 \\ -0{,}2 & 1 & 0 \\ -0{,}6 & 5{,}5 & 1 \end{pmatrix}$, $U = \begin{pmatrix} -5 & 2 & -1 \\ 0 & 0{,}4 & 2{,}8 \\ 0 & 0 & -10 \end{pmatrix}$)

10. Choleského metodou řešte soustavu

$$\begin{aligned} x_1 + x_2 + x_3 &= 2 \\ x_1 + 5x_2 + 5x_3 &= 5 \\ x_1 + 5x_2 + 14x_3 &= 8 \end{aligned}$$

11. Choleského metodou řešte systém

$$\begin{aligned} x_1 + 3x_2 - 2x_3 - 2x_5 &= 0{,}5 \\ 3x_1 + 4x_2 - 5x_3 + x_4 - 3x_5 &= 5{,}4 \\ -2x_1 - 5x_2 + 3x_3 - 2x_4 + 2x_5 &= 5{,}0 \\ x_2 - 2x_3 + 5x_4 + 3x_5 &= 7{,}5 \\ -2x_1 - 3x_2 + 2x_3 + 3x_4 + 4x_5 &= 3{,}3. \end{aligned}$$

(Řešení: $\boldsymbol{x}^* = (-0{,}60978; -2{,}2016; -6{,}8011; -0{,}8996; 0{,}1995)^T$.)

12. Croutovou metodou řešte systém

$$\begin{aligned} 4x_1 + 3x_2 &= 24 \\ 3x_1 + 4x_2 - x_3 &= 30 \\ -x_2 + 4x_3 &= -24. \end{aligned}$$

(Řešení: $\boldsymbol{x}^* = (3, 4, -5)^T$.)

13. Nechť A je pozitivně definitní matice. Ukažte, že

    a) $a_{ii} > 0$, $i = 1, 2, \ldots, n$,

    b) $\max\limits_{1 \le i \le n} a_{ii} = \max\limits_{i,j} |a_{ij}|$.

Kontrolní otázky ke kapitole 4

1. Je možné provést rozklad $A = LR$, respektive $PA = LR$ pro singulární matici A?

2. Popište, jak byste pomocí GEM řešili tuto úlohu:
   Je dáno m systémů lineárních rovnic vždy s toutéž maticí A. Tato úloha může být zapsána ve tvaru
   $$AX = B\,,$$
   $A \in \mathcal{M}_n$, $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$, $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ jsou matice typu $n \times m$, $\boldsymbol{b}_i$, $\boldsymbol{b}_i$, $\boldsymbol{x}_i$, $i = 1, \ldots, m$ jsou vektory.

3. Lze užít elementární matici E_1 definovanou vztahem ($A_{11} \neq 0$)
   $$E_1 = \begin{pmatrix} 1 & & & & \\ -\frac{a_{21}}{a_{11}} & \ddots & & 0 & \\ \vdots & & 1 & & \\ \vdots & & 0 & \ddots & \\ -\frac{a_{n1}}{a_{11}} & & & & 1 \end{pmatrix}$$
   pro transformaci matice $A = A^{(1)}$ na matici $A^{(2)}$ v Gaussově eliminační metodě?

4. Lze použít Choleského metodu pro řešení systému s maticí
   $$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ?$$

5. Je možné rozložit matici
   $$A = \begin{pmatrix} 3 & 3 & 2 \\ -1 & -1 & 4 \\ 2 & 8 & -2 \end{pmatrix}$$
   na součin dolní a horní trojúhelníkové matice?

Kapitola 5

Iterační metody řešení systémů lineárních rovnic

Uvažujme systém lineárních rovnic

$$A\boldsymbol{x} = \boldsymbol{b} \tag{5.1}$$

s regulární maticí $A \in \mathcal{M}_n$. Označme, stejně jako v předchozí kapitole, $\boldsymbol{x}^*$ přesné řešení tohoto systému,

$$\boldsymbol{x}^* = A^{-1}\boldsymbol{b}. \tag{5.2}$$

Přímé metody založené na rozkladu matice A nejsou vhodné vzhledem k době výpočtu a nárokům na paměť počítače v případě, že matice A je dosti velká. V praxi se s takovými maticemi setkáváme např. při numerickém řešení parciálních diferenciálních rovnic, kdy se často vyskytují matice řádu $n > 10000$. V těchto případech je použití Gaussovy eliminační metody velmi náročné. Na druhé straně tyto matice jsou často řídké, tj. mají velké procento nulových prvků, ale tato vlastnost se „ztrácí“ užitím metod předchozí kapitoly. Pro řešení takových úloh lze užít třídy metod, které se nazývají *iterační metody*. Tyto metody nemění strukturu matice A a požadují uchování pouze několika vektorů řádu n.

§ 5.1. Princip iteračních metod

Základní myšlenka iteračních metod spočívá nejdříve ve vyjádření systému $A\boldsymbol{x} = \boldsymbol{b}$ v ekvivalentním tvaru

$$\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{g}, \qquad T \in \mathcal{M}_n, \tag{5.3}$$

$\boldsymbol{x}^*$ je řešení systému (5.1) právě tehdy, když $\boldsymbol{x}^*$ je řešením systému (5.3), $\boldsymbol{x}^* = (E - T)^{-1}\boldsymbol{g}$ za předpokladu, že $E - T$ je regulární.

Nechť $\boldsymbol{x}^0 \in \mathbb{R}^n$ je libovolná počáteční aproximace. Posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ určená rekurentně vztahem

$$\boldsymbol{x}^{k+1} = T\boldsymbol{x}^k + \boldsymbol{g}, \qquad k = 0, 1, \ldots \tag{5.4}$$

se nazývá *iterační posloupnost* a matice T se nazývá *iterační matice*.

Budeme se nyní zabývat dvěma problémy:

(a) Jak zvolit iterační matici T, tj. jakým způsobem převést systém (5.1) na systém (5.3)?

(b) Za jakých předpokladů posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ konverguje pro libovolnou počáteční aproximaci k přesnému řešení $\boldsymbol{x}^*$?

Všimněme si, že vztah (5.4) může být také zapsán jiným způsobem:

Je

$$\begin{aligned} \boldsymbol{x}^1 &= T\boldsymbol{x}^0 + \boldsymbol{g}, \\ \boldsymbol{x}^2 &= T\boldsymbol{x}^1 + \boldsymbol{g} = T(T\boldsymbol{x}^0 + \boldsymbol{g}) + \boldsymbol{g} = T^2\boldsymbol{x}^0 + (T+E)\boldsymbol{g}, \\ &\vdots \\ \boldsymbol{x}^{k+1} &= T^{k+1}\boldsymbol{x}^0 + (T^k + T^{k-1} + \ldots + E)\boldsymbol{g}. \end{aligned} \tag{5.5}$$

Mocniny matice T budou hrát zřejmě důležitou úlohu v našich dalších úvahách. Podívejme se na posloupnosti mocnin matic obecně.

Definice 5.1. Řekneme, že matice $H \in \mathcal{M}_n$ je *konvergentní*, jestliže

$$\lim_{k\to\infty} H^k = \lim_{k\to\infty} \underbrace{H \ldots H}_{k\text{-krát}} = O,$$

kde O je nulová matice.

Příklad 5.1. Matice $H = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$ je konvergentní. Je totiž

$$H^2 = \begin{pmatrix} \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}, \quad H^3 = \begin{pmatrix} \frac{1}{8} & 0 \\ \frac{3}{16} & \frac{1}{8} \end{pmatrix}, \quad \ldots, \quad H^k = \begin{pmatrix} \left(\frac{1}{2}\right)^k & 0 \\ \frac{k}{2^{k+1}} & \left(\frac{1}{2}\right)^k \end{pmatrix}$$

a odtud je zřejmé, že $\lim_{k\to\infty} H^k = O$.

Věta 5.1. ([8]) *Následující tvrzení jsou ekvivalentní:*

(i) *H je konvergentní matice.*

(ii) $\lim_{k\to\infty} \|H^k\| = 0$ *pro nějakou přidruženou maticovou normu.*

(iii) $\varrho(H) < 1$ *($\varrho(H)$ je spektrální poloměr H).*

(iv) $\lim_{k\to\infty} H^k\boldsymbol{x} = \boldsymbol{o}$ *pro libovolný vektor* $\boldsymbol{x} \in \mathbb{R}^n$.

Příklad 5.2. Matice $H = \begin{pmatrix} 1 & 0 \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$ není konvergentní, neboť $\varrho(H) = 1$.

V našich dalších úvahách budeme používat poznatků o normách vektorů a matic z kapitoly 1.

Vraťme se nyní k iteračnímu procesu (5.4). Dříve než dokážeme hlavní větu o konvergenci iteračního procesu, dokážeme toto lemma:

Lemma. *Nechť $\varrho(T) < 1$. Pak $E - T$ je regulární a platí*

$$(E - T)^{-1} = E + T + T^2 + \dots \tag{5.6}$$

Důkaz. První část tvrzení plyne z důsledku věty 1.5. Dokážeme platnost (5.6). Nechť

$$S_m = E + T + T^2 + \dots + T^m.$$

Pak

$$(E - T)S_m = (E - T)(E + \dots + T^m) = E - T^{m+1}.$$

Jelikož $\varrho(T) < 1$, je matice T konvergentní a tudíž $\lim\limits_{m\to\infty} T^{m+1} = O$. Odtud

$$\lim_{m\to\infty} (E - T)S_m = E,$$

což znamená, že

$$(E - T)^{-1} = E + T + T^2 + \dots$$

□

Hlavní větu o konvergenci iteračního procesu (5.4) lze formulovat takto:

Věta 5.2. *Posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ určená iteračním procesem (5.4) konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$ právě tehdy, když $\varrho(T) < 1$, přičemž $\lim\limits_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{x}^*$, $\boldsymbol{x}^* = T\boldsymbol{x}^* + \boldsymbol{g}$.*

Důkaz. Nechť $\boldsymbol{x}^0 \in \mathbb{R}^n$ je libovolná počáteční aproximace. Podle vztahu (5.5) lze aproximaci $\boldsymbol{x}^{k+1}$ zapsat ve tvaru:

$$\boldsymbol{x}^{k+1} = T^{k+1}\boldsymbol{x}^0 + (T^k + T^{k-1} + \dots + T + E)\boldsymbol{g}.$$

Nechť $\varrho(T) < 1$. Pak podle věty 5.1 je matice T konvergentní a podle lemmatu je $(E - T)$ regulární. Odtud plyne, že

$$\lim_{k\to\infty} \boldsymbol{x}^{k+1} = \lim_{k\to\infty} T^{k+1}\boldsymbol{x}^0 + \lim_{k\to\infty} (T^k + \dots + T + E)\boldsymbol{g} = \boldsymbol{o} + (E - T)^{-1}\boldsymbol{g} = \boldsymbol{x}^*.$$

Nechť nyní iterační proces (5.4) konverguje k limitě $\boldsymbol{x}^*$ pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Nechť $\boldsymbol{x}^k = T\boldsymbol{x}^{k-1} + \boldsymbol{g}$, $k = 1, 2, \dots$, $\boldsymbol{x}^* = T\boldsymbol{x}^* + \boldsymbol{g}$. Pak

$$\boldsymbol{x}^* - \boldsymbol{x}^k = T(\boldsymbol{x}^* - \boldsymbol{x}^{k-1}) = \dots = T^k(\boldsymbol{x}^* - \boldsymbol{x}^0). \tag{5.7}$$

Odtud pro libovolný vektor $\boldsymbol{x}^0 \in \mathbb{R}^n$ platí

$$\lim_{k\to\infty}(\boldsymbol{x}^* - \boldsymbol{x}^k) = \lim_{k\to\infty} T^k(\boldsymbol{x}^* - \boldsymbol{x}^0) = \boldsymbol{o}.$$

Nechť nyní $\boldsymbol{z} \in \mathbb{R}^n$ je libovolný vektor a položme $\boldsymbol{x}^0 = \boldsymbol{x}^* - \boldsymbol{z}$, pak

$$\lim_{k\to\infty} T^k\boldsymbol{z} = \lim_{k\to\infty} T^k(\boldsymbol{x}^* - (\boldsymbol{x}^* - \boldsymbol{z})) = \boldsymbol{o},$$

což implikuje, podle věty 5.1, že $\varrho(T) < 1$. □

Poznámka 1. Kriteria pro zastavení výpočtu mohou být např. následující:

1. $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|/\|\boldsymbol{x}^k\| < \varepsilon$, kde $\|\ .\ \|$ je nějaká vektorová norma a $\varepsilon > 0$ je požadovaná přesnost,
2. $\|\boldsymbol{r}^{k+1}\| \le \varepsilon(\|A\|\,\|\boldsymbol{x}^{k+1}\| + \|\boldsymbol{b}\|)$, kde $\boldsymbol{r}^{k+1} = A\boldsymbol{x}^{k+1} - \boldsymbol{b}$,

maticová norma je přidružená dané vektorové normě, a $\varepsilon > 0$ je požadovaná přesnost.

Víme, že pro přidruženou maticovou normu platí $\|T\| \ge \varrho(T)$. Nutnou a postačující podmínku $\varrho(T) < 1$ lze pak ve větě 5.2 nahradit podmínkou postačující: $\|T\| < 1$. Tuto skutečnost zformulujme jako důsledek.

Důsledek. *Nechť pro nějakou přidruženou maticovou normu platí $\|T\| < 1$. Pak posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ generovaná iteračním procesem (5.4) konverguje k řešení $\boldsymbol{x}^* = (E - T)^{-1}\boldsymbol{g}$ pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$. Dále platí*

$$\|\boldsymbol{x}^* - \boldsymbol{x}^k\| \le \|T\|^k\|\boldsymbol{x}^* - \boldsymbol{x}^0\|, \tag{5.8}$$

$$\|\boldsymbol{x}^* - \boldsymbol{x}^k\| \le \frac{\|T\|^k}{1 - \|T\|}\|\boldsymbol{x}^1 - \boldsymbol{x}^0\|. \tag{5.9}$$

Důkaz. Jak již bylo uvedeno, první část důkazu plyne ze skutečnosti $1 > \|T\| \ge \varrho(T)$ a je tedy důsledkem předchozí věty. Ale důkaz lze rovněž provést aplikací Banachovy věty o pevném bodě:

Uvažujme zobrazení $F\colon\ \mathbb{R}^n \to \mathbb{R}^n$, $F\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{g}$. Ukážeme, že toto zobrazení je kontrakce v prostoru $\mathbb{R}^n$ vzhledem k metrice $\varrho(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$, $\mathbb{R}^n$ je úplný metrický prostor. Je totiž

$$\|F\boldsymbol{x} - F\boldsymbol{y}\| = \|T(\boldsymbol{x} - \boldsymbol{y})\| \le \|T\|\,\|\boldsymbol{x} - \boldsymbol{y}\| = q\|\boldsymbol{x} - \boldsymbol{y}\|.$$

To znamená, že F je kontrakce s koeficientem $q = \|T\|$. Z Banachovy věty o pevném bodě plyne tvrzení a rovněž vztah (5.9).

Pokud jde o odhad daný vztahem (5.8), plyne ihned ze vztahu (5.7), neboť

$$\|\boldsymbol{x}^* - \boldsymbol{x}^k\| = \|T^k(\boldsymbol{x}^* - \boldsymbol{x}^0)\| \le \|T\|^k\|\boldsymbol{x}^* - \boldsymbol{x}^0\|.$$

□

Připomeňme ještě zajímavý výsledek, týkající se vztahu přidružené normy matice a jejího spektrálního poloměru: Pro každou matici A a libovolné $\varepsilon > 0$ existuje

přidružená maticová norma s vlastností $\|A\| \leq \varrho(A) + \varepsilon$. Tedy $\varrho(A)$ je infimum všech přidružených norem matice A ([17]).

Vraťme se nyní k iteračním procesům. Jelikož vztah (5.8) platí pro každou přidruženou maticovou normu, plyne z předchozího

$$\|\boldsymbol{x}^k - \boldsymbol{x}^*\| \approx (\varrho(T))^k \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|. \tag{5.10}$$

Předpokládejme, že $\varrho(T) < 1$ a $\boldsymbol{x}^0 = \boldsymbol{o}$ je počáteční aproximace. Chceme-li dosáhnout relativní chyby nejvýše 10^{-t}, je podle vztahu (5.10) zapotřebí k iterací, přičemž pro k platí

$$(\varrho(T))^k \leq 10^{-t},$$

tj.

$$k \geq -\frac{t}{\log \varrho(T)}. \tag{5.11}$$

§ 5.2. Jacobiova iterační metoda

Volbou iterační matice T lze získat konkrétní iterační metody.

Matici A zapišme ve tvaru

$$A = D - L - U,$$

kde

$$D = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix},$$

$$L = \begin{pmatrix} 0 & & & 0 \\ -a_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ -a_{n1} & \cdots & -a_{n,n-1} & 0 \end{pmatrix},$$

$$U = \begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & -a_{n-1,n} \\ 0 & & & 0 \end{pmatrix}.$$

D je diagonální matice, L je dolní trojúhelníková matice s nulami na diagonále a U je horní trojúhelníková matice s nulami na diagonále.

Rovnici $A\boldsymbol{x} = \boldsymbol{b}$ zapíšeme ve tvaru $(D - L - U)\boldsymbol{x} = \boldsymbol{b}$ a transformujeme ji na rovnici

$$D\boldsymbol{x} = (L + U)\boldsymbol{x} + \boldsymbol{b}.$$

Za předpokladu, že $a_{ii} \neq 0$, $i = 1, \ldots, n$, je matice D regulární a z předchozí rovnice lze vypočítat $\boldsymbol{x}$

$$\boldsymbol{x} = D^{-1}(L + U)\boldsymbol{x} + D^{-1}\boldsymbol{b}. \tag{5.12}$$

Tento vztah vede na maticový tvar *Jacobiovy iterační metody*. Označíme-li $T_J = D^{-1}(L + U)$, je tato metoda tvaru:

$$\boldsymbol{x}^{k+1} = T_J\boldsymbol{x}^k + D^{-1}\boldsymbol{b}, \tag{5.13}$$

kde $T_J = (t_{ij})$ je *Jacobiova iterační matice*, $t_{ij} = -\frac{a_{ij}}{a_{ii}}$ pro $i \neq j$, $t_{ii} = 0$ pro $i = 1, \ldots, n$. Matice T_J má tedy nulové diagonální prvky a je tvaru

$$T_J = \begin{pmatrix} 0 & -\frac{a_{12}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & & -\frac{a_{2n}}{a_{22}} \\ \vdots & & \ddots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \cdots & 0 \end{pmatrix}. \tag{5.14}$$

Ve složkách vektoru $\boldsymbol{x}^k$ lze Jacobiovu iterační metodu zapsat takto:

$$x_i^{k+1} = -\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}}{a_{ii}}x_j^k + \frac{b_i}{a_{ii}}, \qquad i = 1, \ldots, n;\ k \geq 0. \tag{5.15}$$

Pro realizaci výpočtu to znamená, že *z první rovnice vypočteme* x_1, *z druhé* x_2, *obecně z* k*-té rovnice vypočteme* x_k *až z* n*-té rovnice vypočteme* x_n a na pravé straně takto získaného systému jsou prvky matice T_J.

Z věty 5.2 ihned plyne věta o konvergenci Jacobiovy iterační metody:

Věta 5.3. *Posloupnost* $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ *generovaná metodou* (5.13) *konverguje pro každou počáteční aproximaci* $\boldsymbol{x}^0 \in \mathbb{R}^n$ *právě tehdy, když* $\varrho(T_J) < 1$.

Příklad 5.3. Jacobiovou iterační metodou řešte systém

$$\begin{aligned} 10x_1 - 2x_2 - 2x_3 &= 6 \\ -x_1 + 10x_2 - 2x_3 &= 7 \\ -x_1 - x_2 + 10x_3 &= 8, \end{aligned}$$

jehož přesné řešení $\boldsymbol{x}^* = (1, 1, 1)^T$.

Řešení. Jacobiova iterační metoda je tvaru

$$\begin{aligned} x_1^{k+1} &= \tfrac{1}{10}(6 + 2x_2^k + 2x_3^k) \\ x_2^{k+1} &= \tfrac{1}{10}(7 + x_1^k + 2x_3^k) \\ x_3^{k+1} &= \tfrac{1}{10}(8 + x_1^k + x_2^k) \end{aligned}$$

Nechť $\boldsymbol{x}^0 = (0,0,0)^T$. Pak

$$\boldsymbol{x}^1 = \begin{pmatrix} 0{,}6 \\ 0{,}7 \\ 0{,}8 \end{pmatrix} \quad \boldsymbol{x}^2 = \begin{pmatrix} 0{,}90 \\ 0{,}92 \\ 0{,}93 \end{pmatrix} \quad \boldsymbol{x}^3 = \begin{pmatrix} 0{,}970 \\ 0{,}976 \\ 0{,}982 \end{pmatrix} \quad \boldsymbol{x}^4 = \begin{pmatrix} 0{,}9918 \\ 0{,}9934 \\ 0{,}9958 \end{pmatrix}$$

Matice T_J je v tomto případě tvaru

$$T_J = \begin{pmatrix} 0 & 0{,}2 & 0{,}2 \\ 0{,}1 & 0 & 0{,}2 \\ 0{,}1 & 0{,}1 & 0 \end{pmatrix} \qquad \varrho(T_J) = 0{,}285$$

a $\|T_J\|_\infty = 0{,}4$. Podle věty 5.3 je posloupnost $\{\boldsymbol{x}^k\}$ konvergentní.

Pro odhad chyby platí (viz (5.9))

$$\|\boldsymbol{x}^* - \boldsymbol{x}^k\|_\infty \leq \frac{\|T_J\|_\infty^k}{1 - \|T\|_\infty} \|\boldsymbol{x}^1 - \boldsymbol{x}^0\|_\infty,$$

což v našem případě pro $k = 4$

$$\|\boldsymbol{x}^* - \boldsymbol{x}^4\|_\infty \leq \frac{0{,}4^4}{0{,}6} 0{,}8 \approx 0{,}034.$$

Na druhé straně, skutečná chyba $\|\boldsymbol{x}^* - \boldsymbol{x}^4\|_\infty \approx 0{,}008$. Vztah (5.9) udává totiž, jako obvykle v numerických metodách, horní odhad chyby.

Pro některé speciální typy matice A je zaručena konvergence Jacobiovy iterační metody.

Věta 5.4.

a) Silné řádkové sumační kriterium*:*

Nechť matice A je ryze řádkově diagonálně dominantní, tj.

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \qquad i = 1, \ldots, n.$$

Pak Jacobiova iterační metoda konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

b) Silné sloupcové sumační kriterium*:*

Nechť matice A je ryze sloupcově diagonálně dominantní, tj.

$$|a_{kk}| > \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ik}|, \qquad k = 1, \ldots, n.$$

Pak Jacobiova iterační metoda konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Důkaz.

a) Spočítejme normu $\| \, . \, \|_\infty$ matice $T_J = D^{-1}(L+U)$. Je

$$\|T_J\|_\infty = \max_{1\le i \le n} \sum_{j=1}^{n} |t_{ij}| = \max_{1\le i \le n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left|\frac{a_{ij}}{a_{ii}}\right|.$$

Jelikož

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|,$$

je

$$\sum_{j=1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| < 1, \qquad i = 1, \ldots, n$$

a odtud $\|T_J\|_\infty < 1$. Podle důsledku věty 5.2 je Jacobiova iterační metoda konvergentní.

b) Podle předpokladu Jacobiova metoda konverguje pro matici A^T. To znamená, že $\varrho(D^{-1}(L^T + U^T)) < 1$. Položme $X = D^{-1}(L^T + U^T)$. Tato matice má stejná vlastní čísla jako matice $X^T = (L+U)D^{-1}$. Dále matice X^T má stejná vlastní čísla jako matice s ní podobná

$$D^{-1}X^T D = D^{-1}(L+U)D^{-1}D = D^{-1}(L+U) = T_J$$

a odtud plyne, že $\varrho(T_J) < 1$ a je splněn předpoklad věty 5.3.

□

Matice A systému příkladu 5.3 je ryze řádkově i sloupcově diagonálně dominantní.

Geometrický význam Jacobiovy metody budeme ilustrovat na příkladu systému dvou rovnic:

Příklad 5.4. Uvažujme systém

$$\begin{aligned} l_1\colon &\quad a_{11}x_1 + a_{12}x_2 = b_1 \\ l_2\colon &\quad a_{21}x_1 + a_{22}x_2 = b_2. \end{aligned}$$

Tyto rovnice jsou rovnice přímek l_1, l_2. Jacobiova iterační metoda je tvaru:

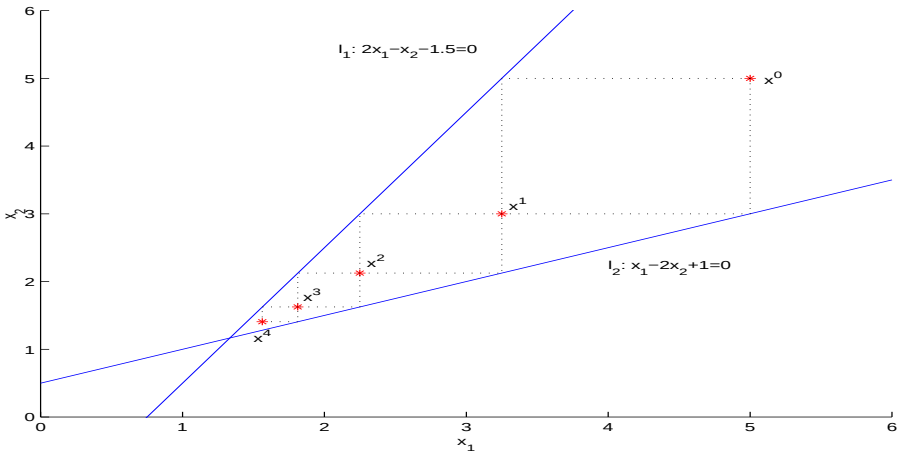
$$\begin{aligned} l_1\colon &\quad x_1^{k+1} = -\frac{a_{12}}{a_{11}}x_2^k + \frac{b_1}{a_{11}} \\ l_2\colon &\quad x_2^{k+1} = -\frac{a_{21}}{a_{22}}x_1^k + \frac{b_2}{a_{22}} \end{aligned} \tag{5.16}$$

a iterační matice

$$T_J = \begin{pmatrix} 0 & -\dfrac{a_{12}}{a_{11}} \\ -\dfrac{a_{21}}{a_{22}} & 0 \end{pmatrix}$$

má charakteristickou rovnici $\lambda^2 + a_{12}a_{21}/a_{11}a_{22} = 0$.

Ze vztahů (5.16) bezprostředně plyne geometrický význam této metody:
Bod (x_1^{k+1}, x_2^k) leží na přímce l_1, bod (x_1^k, x_2^{k+1}) leží na přímce l_2. Bod (x_1^{k+1}, x_2^{k+1}) je průsečíkem přímek $x_2 = x_1^{k+1}$ a $x_2 = x_2^{k+1}$.
Graficky je tato metoda ilustrována na obr. 5.1.



Obr. 5.1: Jacobiova iterační metoda

Poznámka 2. Vlastní čísla matice T_J jsou dána vztahem $\lambda^2 = -a_{12}a_{21}/a_{11}a_{22}$. Vyměníme-li pořadí přímek (za předpokladu, že $a_{ij} \neq 0$, $i, j = 1, 2$) budou vlastní čísla tohoto systému rovna převráceným hodnotám čísel λ, což má za následek změnu konvergentního procesu v proces divergentní a naopak.

§ 5.3. Gaussova-Seidelova iterační metoda

Při použití Jacobiovy metody při výpočtu $\boldsymbol{x}^{k+1}$ musíme uchovávat v paměti počítače celý vektor $\boldsymbol{x}^k$. Jistou modifikací Jacobiovy metody je metoda, která při výpočtu složky x_i^{k+1}, $1 < i \leq n$, používá již vypočtené složky $x_1^{k+1}, \ldots, x_{i-1}^{k+1}$. Popišme nyní tuto metodu podrobněji. Uvedeme nejdříve zápis po složkách a poté přejdeme k maticovému zápisu.

Uvažovaná metoda může být zapsána takto:

$$\begin{aligned}
a_{11}x_1^{k+1} &+ a_{12}x_2^k &&+ \ldots + a_{1n}x_n^k &&= b_1\\
a_{21}x_1^{k+1} &+ a_{22}x_2^{k+1} &&+ \ldots + a_{2n}x_n^k &&= b_2\\
\vdots & && && \vdots\\
a_{n1}x_1^{k+1} &+ a_{n2}x_2^{k+1} &&+ \ldots + a_{nn}x_n^{k+1} &&= b_n
\end{aligned}$$

neboli

$$x_i^{k+1} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{k+1} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^k + \frac{b_i}{a_{ii}}, \qquad i = 1, \ldots, n. \tag{5.17}$$

Tato metoda se nazývá *Gaussova-Seidelova metoda.*

Vztah (5.17), který je v podstatě získaný obdobným postupem jako u Jacobiovy metody, je ***vhodný pro praktické použití.***

Maticový zápis Gaussovy-Seidelovy metody lze získat takto:

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} \quad \Rightarrow \quad (D - L - U)\boldsymbol{x} &= \boldsymbol{b} \\ (D - L)\boldsymbol{x} &= U\boldsymbol{x} + \boldsymbol{b}. \end{aligned}$$

Za předpokladu, že $a_{ii} \neq 0$, $i = 1, \ldots, n$, je matice $D - L$ regulární a

$$\boldsymbol{x} = (D - L)^{-1} U\boldsymbol{x} + (D - L)^{-1}\boldsymbol{b}.$$

Položíme-li $T_G = (D - L)^{-1}U$, je Gaussova-Seidelova iterační metoda tvaru

$$\boldsymbol{x}^{k+1} = T_G\boldsymbol{x}^k + \boldsymbol{g}, \qquad \boldsymbol{g} = (D - L)^{-1}\boldsymbol{b}. \tag{5.18}$$

Věta 5.5. *Posloupnost* $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ *generovaná Gaussovou-Seidelovou iterační metodou* (5.18) *konverguje pro každou počáteční aproximaci* $\boldsymbol{x}^0 \in \mathbb{R}^n$ *právě tehdy, když* $\varrho(T_G) < 1$.

Důkaz. Věta je přímým důsledkem věty 5.2. □

Poznámka 3. Vztah (5.18) umožňuje stanovit kritéria pro konvergenci metody.

Věta 5.6. *Nechť jsou splněny předpoklady a), b) věty* 5.4. *Pak Gaussova-Seidelova metoda konverguje pro každou počáteční aproximaci* $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Důkaz.
Důkaz provedeme pouze pro ryze řádkově diagonálně dominantní matice.
Nechť λ je vlastní číslo matice $T_G = (D - L)^{-1}U$ a nechť $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ je odpovídající vlastní vektor. Odtud plyne, že

$$T_G\boldsymbol{x} = \lambda\boldsymbol{x} \quad \Rightarrow \quad (D - L)^{-1}U\boldsymbol{x} = \lambda\boldsymbol{x} \quad \Rightarrow \quad U\boldsymbol{x} = \lambda(D - L)\boldsymbol{x},$$

neboli vyjádřeno ve složkách vektoru $\boldsymbol{x}$

$$-\sum_{j=i+1}^{n} a_{ij}x_j = \lambda \sum_{j=1}^{i} a_{ij}x_j.$$

Poslední vztah můžeme přepsat takto:

$$\lambda a_{ii}x_i = -\lambda \sum_{j=1}^{i-1} a_{ij}x_j - \sum_{j=i+1}^{n} a_{ij}x_j, \qquad 1 \leq i \ \leq n.$$

Nechť $|x_k| = \max\limits_{1\leq i\ \leq n} |x_i|$. Nyní pro $i = k$ dostaneme z předchozího vztahu

$$|\lambda||a_{kk}| \leq |\lambda| \sum_{j=1}^{k-1} |a_{kj}| + \sum_{j=k+1}^{n} |a_{kj}|,$$

což znamená, že

$$|\lambda| \left(|a_{kk}| - \sum_{j=1}^{k-1} |a_{kj}| \right) \leq \sum_{j=k+1}^{n} |a_{kj}|$$

neboli

$$|\lambda| \leq \frac{\sum\limits_{j=k+1}^{n} |a_{kj}|}{\left(|a_{kk}| - \sum\limits_{j=1}^{k-1} |a_{kj}| \right)} \tag{5.19}$$

Matice A je ryze řádkově diagonálně dominantní, tj.

$$|a_{kk}| > \sum_{\substack{j=1 \\ j\neq k}}^{n} |a_{kj}|,$$

což také znamená, že

$$|a_{kk}| - \sum_{j=1}^{k-1} |a_{kj}| > \sum_{j=k+1}^{n} |a_{kj}|. \tag{5.20}$$

Nyní z (5.20) a (5.18) plyne, že $|\lambda| < 1$ a tedy $\varrho(T_G) < 1$ a věta je dokázána. □

Příklad 5.5. Systém v příkladě 5.3 řešte Gaussovou-Seidelovou iterační metodou.
Řešení. Gaussova-Seidelova metoda bude konvergovat pro každou počáteční aproximaci, neboť matice systému je ryze řádkově diagonálně dominantní.

Tato metoda je nyní tvaru

$$\begin{aligned} x_1^{k+1} &= \tfrac{1}{10}(6 + 2x_2^k \quad + 2x_3^k) \\ x_2^{k+1} &= \tfrac{1}{10}(7 + \ x_1^{k+1} + 2x_3^k) \\ x_3^{k+1} &= \tfrac{1}{10}(8 + \ x_1^{k+1} + \ x_2^{k+1}) \end{aligned}$$

Nechť $\boldsymbol{x}^0 = (0,0,0)^T$. Pak

$$\boldsymbol{x}^1 = \begin{pmatrix} 0{,}6 \\ 0{,}76 \\ 0{,}936 \end{pmatrix} \quad \boldsymbol{x}^2 = \begin{pmatrix} 0{,}9392 \\ 0{,}98112 \\ 0{,}99203 \end{pmatrix} \quad \boldsymbol{x}^3 = \begin{pmatrix} 0{,}994630 \\ 0{,}997869 \\ 0{,}9992499 \end{pmatrix}$$

Nyní $\|\boldsymbol{x}^* - \boldsymbol{x}^3\|_\infty \approx 0{,}005$.

Iterační matice T_G je tvaru

$$T_G = \begin{pmatrix} 0 & 0{,}2 & 0{,}2 \\ 0 & 0{,}02 & 0{,}22 \\ 0 & 0{,}022 & 0{,}042 \end{pmatrix} \qquad \varrho(T_G) = 0{,}101.$$

Podívejme se nyní na geometrický význam Gaussovy-Seidelovy metody.

Příklad 5.6. Uvažujme systém dvou rovnic:

$$\begin{aligned} l_1\colon &\quad a_{11}x_1 + a_{12}x_2 = b_1 \\ l_2\colon &\quad a_{21}x_1 + a_{22}x_2 = b_2 \end{aligned}$$

Gaussova-Seidelova iterační metoda je tvaru

$$\begin{aligned} l_1\colon &\quad x_1^{k+1} = -\frac{a_{12}}{a_{11}}x_2^k \quad + \frac{b_1}{a_{11}} \\ l_2\colon &\quad x_2^{k+1} = -\frac{a_{21}}{a_{22}}x_1^{k+1} + \frac{b_2}{a_{22}} \end{aligned} \tag{5.21}$$

Iterační matici T_G lze v tomto případě jednoduše získat dosazením x_1^{k+1} z první rovnice do druhé rovnice:

$$\begin{aligned} l_1\colon &\quad x_1^{k+1} = \quad -\frac{a_{12}}{a_{11}}x_2^k + \frac{b_1}{a_{11}} \\ l_2\colon &\quad x_2^{k+1} = -\frac{a_{12}a_{21}}{a_{11}a_{22}}x_2^k + \frac{b_2}{a_{22}} - \frac{a_{21}b_1}{a_{22}a_{11}} \end{aligned}$$

Iterační matice T_G

$$T_G = \begin{pmatrix} 0 & -\dfrac{a_{12}}{a_{11}} \\ 0 & \dfrac{a_{12}a_{21}}{a_{11}a_{22}} \end{pmatrix}$$

je singulární. Vlastní čísla této matice vyhovují rovnici

$$\lambda\left(\lambda - \frac{a_{12}a_{21}}{a_{11}a_{22}}\right) = 0,$$

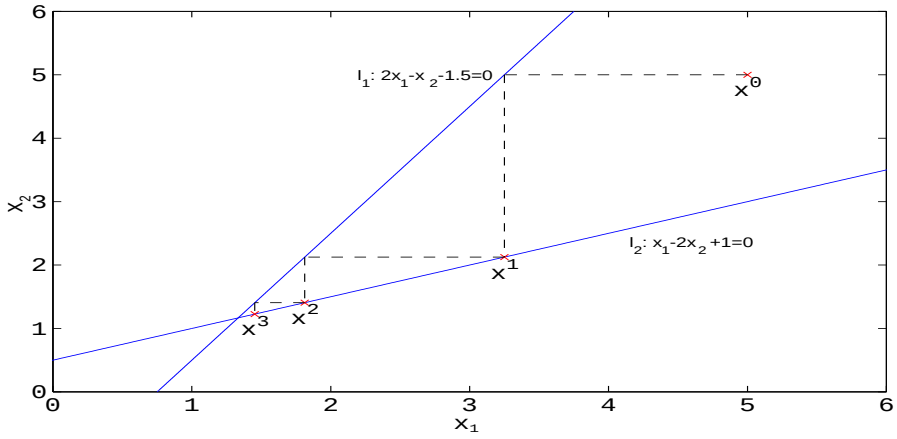
tj. $\lambda_1 = 0$, $\lambda_2 = a_{12}a_{21}/a_{11}a_{22}$.

Z geometrického hlediska bod o souřadnicích (x_1^{k+1}, x_2^k) leží na přímce l_1 a bod (x_1^{k+1}, x_2^{k+1}) leží na přímce l_2. Tato skutečnost je zřejmá ze vztahů (5.21).

Graficky je iterační proces znázorněn na obr. 5.2.

I v tomto případě, při změně pořadí rovnic, konvergentní proces se změní na divergentní a naopak.

Viděli jsme, že pro ryze řádkově diagonálně dominantní matice Jacobiova i Gaussova-Seidelova metoda konvergují. Přirozeně vzniká otázka, zda toto platí



Obr. 5.2: Gaussova-Seidelova iterační metoda

i pro jiné typy matic a jestliže ano, která z metod konverguje rychleji. Ale obecně jsou obory konvergence těchto dvou metod různé a jen částečně se překrývají. Porovnat obory konvergence těchto metod lze pouze ve speciálních případech.

Věta 5.7. (Stein-Rosenberg). ([5]) *Nechť pro prvky matice A platí $a_{ij} \leq 0$ pro všechna $i \neq j$ a $a_{ii} > 0$, $i = 1, \ldots, n$. Pak platí právě jedno z následujících tvrzení:*

(a) $0 < \varrho(T_G) < \varrho(T_J) < 1$

(b) $1 < \varrho(T_J) < \varrho(T_G)$

(c) $\varrho(T_J) = \varrho(T_G) = 0$

(d) $\varrho(T_J) = \varrho(T_G) = 1$.

To znamená, že konvergují-li obě metody, Gaussova-Seidelova metoda konverguje rychleji.

Podle této věty pro systémy v příkladě 5.5 konverguje Gaussova-Seidelova metoda rychleji než Jacobiova, neboť $\varrho(T_G) < \varrho(T_J)$. Uvedeme ještě jednu větu týkající se Gaussovy-Seidelovy iterační metody.

Věta 5.8. *Nechť A je pozitivně definitní matice. Pak Gaussova-Seidelova metoda konverguje pro každou počáteční aproximaci.*

Důkaz viz [5].

§ 5.4. Relaxační metody

Na základě předchozích výsledků lze vyslovit hypotézu, že existují jednoduché matice T, pro které odpovídající iterační proces konverguje rychleji než v případě

Gaussovy-Seidelovy metody. Uvažujme třídu matic T_ω závisejících na parametru ω a budeme se snažit vybrat parametr ω optimálním způsobem, tj. tak, aby číslo $\varrho(T_\omega)$ bylo co nejmenší. Speciálně uvažujme tuto třídu matic:

$$T_\omega = (D - \omega L)^{-1}[(1-\omega)D + \omega U]$$

Parametr ω se nazývá *relaxační parametr* a odpovídající metody se nazývají *relaxační metody*.

Pro $0 < \omega < 1$ se iterační metody nazývají *metodami dolní relaxace*. Tyto metody jsou vhodné v případě, že Gaussova-Seidelova metoda nekonverguje.

Pro $\omega = 1$ je relaxační metoda totožná s Gaussovou-Seidelovou metodou.

Pro $1 < \omega$ se metody nazývají *metodami horní relaxace*, nebo častěji SOR metodami (SOR = Successive Over-Relaxation). Tyto metody lze užít ke zrychlení konvergence Gaussovy-Seidelovy metody.

Relaxační metodu lze maticově zapsat takto

$$\boldsymbol{x}^{k+1} = (D - \omega L)^{-1}[(1-\omega)D + \omega U]\boldsymbol{x}^k + \omega(D - \omega L)^{-1}\boldsymbol{b} \tag{5.22}$$

a zápis v jednotlivých složkách je tvaru

$$x_i^{k+1} = (1-\omega)x_i^k + \frac{\omega}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{k+1} - \sum_{j=i+1}^{n} a_{ij}x_j^k\right]. \tag{5.23}$$

Je přirozené zabývat se otázkou, pro které hodnoty parametru ω bude iterační proces (5.22) konvergovat. Další otázkou je optimální volba parametru ω. Na první otázku obecně neexistuje vyčerpávající odpověď, ale následující výsledek je velmi důležitý.

Věta 5.9. (Kahan). *Nechť $a_{ii} \neq 0$, $i = 1, \ldots, n$. Pak*

$$\varrho(T_\omega) \geq |\omega - 1|.$$

Důkaz. Připomeňme, že $D - \omega L$ je dolní trojúhelníková matice tvaru

$$D - \omega L = \begin{pmatrix} a_{11} & & & 0 \\ a_{21}\omega & \ddots & & \\ \vdots & & \ddots & \\ a_{n1}\omega & \cdots & \cdots & a_{nn} \end{pmatrix}$$

Je zřejmé, že $\det D = \det(D - \omega L) = \prod_{i=1}^{n} a_{ii}$.

Zabývejme se nyní charakteristickým polynomem $\varphi(\lambda)$ matice T_ω:

$$\begin{aligned}
\varphi(\lambda) &= \det(T_\omega - \lambda E) = \\
&= \det D^{-1} \det(D - \omega L) \det(T_\omega - \lambda E) = \\
&= \det D^{-1} \det(D - \omega L)(T_\omega - \lambda E) = \\
&= \det D^{-1}(D - \omega L) \left((D - \omega L)^{-1}[(1-\omega)D + \omega U] - \lambda E\right) = \\
&= \det D^{-1} \left((1-\omega)D + \omega U - (D - \omega L)\lambda E\right) = \\
&= \det \left((1 - \omega - \lambda)E + \omega D^{-1}(U + \lambda L)\right).
\end{aligned}$$

$\varphi(\lambda)$ je charakteristický polynom; jeho hodnota v bodě $\lambda = 0$ je rovna determinantu dané matice T_ω. Z vlastností kořenů charakteristického polynomu a z předchozího vztahu však dále plyne, že

$$\varphi(0) = \prod_{i=1}^{n} \lambda_i = \det \left((1-\omega)E + \omega D^{-1} U\right),$$

kde λ_i, $i = 1, \ldots, n$, jsou vlastní čísla matice T_ω. Matice $(1-\omega)E + \omega D^{-1}U$ je horní trojúhelníková matice s prvky $1 - \omega$ na diagonále. Odtud plyne, že

$$\det \left((1-\omega)E + \omega D^{-1} U\right) = (1-\omega)^n.$$

Nyní z rovnosti $\prod_{i=1}^{n} \lambda_i = (1-\omega)^n$ nutně plyne, že $\varrho(T_\omega) = \max_{1 \le i \le n} |\lambda_i| \ge |1 - \omega|$. □

Poznámka 4. Z předchozí věty plyne, že má smysl uvažovat pouze $\omega \in (0, 2)$.

Uvažujeme-li nyní stejný systém jako v příkladu 5.4, je příslušná iterační matice T_ω tvaru

$$T_\omega = \begin{pmatrix} 1 - \omega & -\dfrac{a_{12}}{a_{11}}\omega \\ -\dfrac{a_{21}}{a_{22}}\omega(1-\omega) & \dfrac{a_{12}a_{21}}{a_{11}a_{22}}\omega^2 + (1-\omega) \end{pmatrix}$$

a vektor

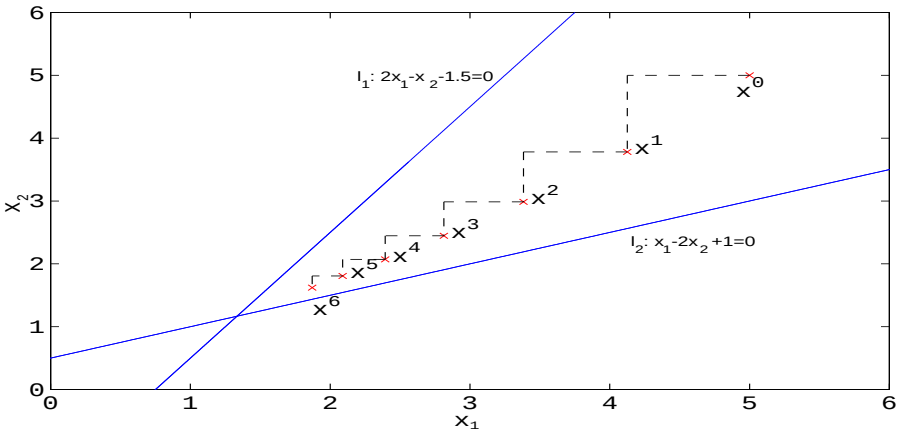
$$\boldsymbol{g}_\omega = \left(\frac{b_1}{a_{11}}\omega, \frac{b_2}{a_{22}}\omega - \frac{b_1 a_{21}}{a_{11}a_{22}}\omega^2\right)^T.$$

Geometricky jsou iterace této metody znázorněny na obr. 5.3 a 5.4.

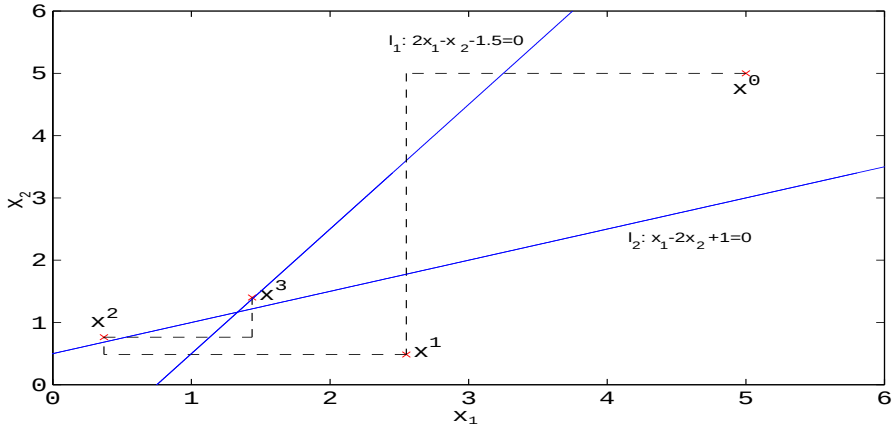
Následující věty udávají postačující podmínky pro konvergenci relaxační metody pro některé typy matic.

Věta 5.10. (Ostrowski-Reich). *Pro pozitivně definitní matici A platí $\varrho(T_\omega) < 1$ pro všechna $\omega \in (0, 2)$.*

Důkaz viz [5].



Obr. 5.3: Relaxační metoda pro $\omega = 0{,}5$



Obr. 5.4: Relaxační metoda pro $\omega = 1{,}4$

Poznámka 5. Z této věty plyne jako důsledek věta 5.8.

Pokud jde o optimální hodnotu parametru ω, tj. hodnotu $\omega \in (0,2)$, pro kterou je konvergence relaxační metody nejrychlejší, uvedeme bez důkazu pouze tuto větu ([5]).

Věta 5.11. *Nechť A je třídiagonální pozitivně definitní matice. Pak $\varrho(T_G) = \varrho^2(T_J) < 1$ a optimální hodnota relaxačního parametru je dána vztahem*

$$\omega = \omega_{opt} = \frac{2}{1 + \sqrt{1 - \varrho^2(T_J)}}.$$

Při této volbě je $\varrho(T_\omega) = |1 - \omega|$.

Příklad 5.7. Uvažujme systém

$$\begin{aligned} 2x_1 - x_3 &= 1 \\ -x_1 + 2x_2 - x_3 &= 0 \\ -x_2 + 2x_3 &= 1, \end{aligned}$$

jehož přesné řešení je $\boldsymbol{x}^* = (1,1,1)^T$.

Matice tohoto systému je třídiagonální a pozitivně definitní a jsou tedy splněny předpoklady věty 5.11. Pro spektrální poloměry iteračních matic platí

$$\varrho(T_G) = \varrho(T_J)^2$$

a optimální hodnota relaxačního parametru ω je dána vztahem

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - (\varrho(T_J))^2}}.$$

Spektrální poloměr iterační matice pro relaxační metodu je roven $\varrho(T_{\omega_{opt}}) = |1 - \omega_{opt}|$.

Jacobiova matice je tvaru

$$T_J = \begin{pmatrix} 0 & 0{,}5 & 0 \\ 0{,}5 & 0 & 0{,}5 \\ 0 & 0{,}5 & 0 \end{pmatrix}.$$

Vlastní čísla této matice jsou $\lambda_1 = 0$, $\lambda_{2,3} = \pm\sqrt{2}/2$. Tedy $\varrho(T_J) = \sqrt{2}/2$, $\varrho(T_G) = 1/2$ a

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - 1/2}} \approx 1{,}172.$$

Za optimální hodnotu parametru ω lze tedy vzít hodnotu 1,17. Pro tuto hodnotu relaxačního parametru je $\varrho(T_{\omega_{opt}}) = 0{,}17$.

Ze vztahu (5.10) plyne, že pro dosažení relativní chyby řádově 10^{-3} je třeba provést 5 iterací, zatímco v případě použití Gaussovy-Seidelovy iterační metody je třeba provést 10 iterací.

Pro daný systém a $\omega = 1{,}17$ je relaxační metoda tvaru

$$\begin{aligned} x_1^{k+1} &= -0{,}17x_1^k + 0{,}585(1 + x_2^k) \\ x_2^{k+1} &= -0{,}17x_2^k + 0{,}585(x_1^{k+1} + x_3^k) \\ x_3^{k+1} &= -0{,}17x_3^k + 0{,}585(1 + x_2^{k+1}) \end{aligned}$$

Pro počáteční aproximaci $\boldsymbol{x}^0 = (0, 0, 0)^T$ dostaneme

$$\begin{array}{lll} x_1^1 = 0{,}585 & x_2^1 = 0{,}3422 & x_3^1 = 0{,}7852 \\ x_1^2 = 0{,}685752 & x_2^2 = 0{,}802329 & x_3^2 = 0{,}920878 \\ x_1^3 = 0{,}9377849 & x_2^3 = 0{,}9509221 & x_3^3 = 0{,}9847401 \\ x_1^4 = 0{,}9818660 & x_2^4 = 0{,}9888078 & x_3^4 = 0{,}9960467 \\ x_1^5 = 0{,}9965353 & x_2^5 = 0{,}9975632 & x_3^5 = 0{,}9992465. \end{array}$$

Z výsledků této kapitoly je zřejmé, že pro $\varrho(T) < 1$ iterační proces konverguje a pro $\varrho(T) > 1$ určitě diverguje.

Povšimněme si nyní chování posloupnosti v případě $\varrho(T) = 1$. Touto otázkou se nebudeme zabývat podrobně, ale upozorníme na zajímavé geometrické chování některých iteračních posloupností.

Definice 5.2. Nechť $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ je posloupnost generovaná iterační metodou (5.4) s počáteční aproximací $\boldsymbol{x}^0 \in \mathbb{R}^n$. Řekneme, že vektor $\boldsymbol{x}^0$ generuje cyklus řádu p, $p \in N$, $p \geq 2$, jestliže $\boldsymbol{x}^p = \boldsymbol{x}^0$, přičemž $\boldsymbol{x}^k \neq \boldsymbol{x}^0$, $k = 1, 2, \ldots, p-1$.

Samozřejmě se předpokládá $\boldsymbol{x}^0 \neq \boldsymbol{x}^*$.

Dá se ukázat (viz [24]), že pro systém tvaru

$$\begin{aligned} x_1 + kx_2 &= b_1 \\ x_1 - kx_2 &= b_2, \end{aligned}$$

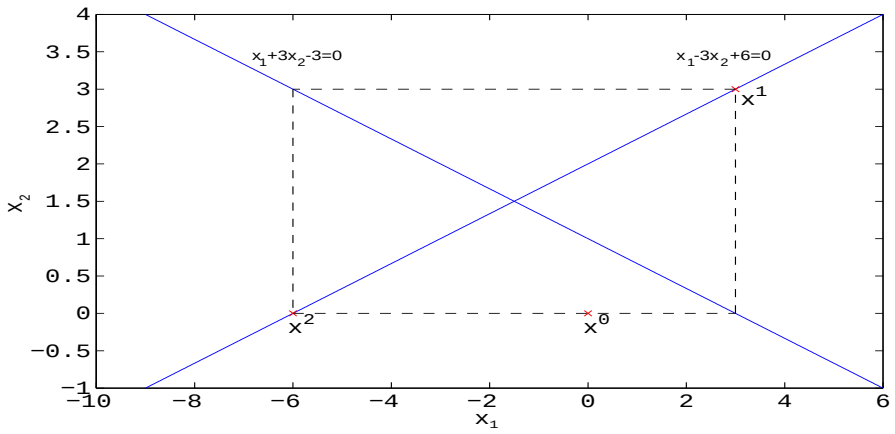
kde $k \neq 0$, nastane pro Jacobiovu metodu cyklus řádu 4 pro každou počáteční aproximaci. V tomto případě $\varrho(T_J) = 1$. Pro Gaussovu-Seidelovu metodu nastane cyklus řádu 2, ale až od 1. aproximace, tj. vlastně s počáteční aproximací Tx^0. Graficky jsou cykly pro tyto metody znázorněny na obr. 5.5, 5.6.

Zajímavá situace se objevuje u relaxační metody. Uvažujme systém tvaru

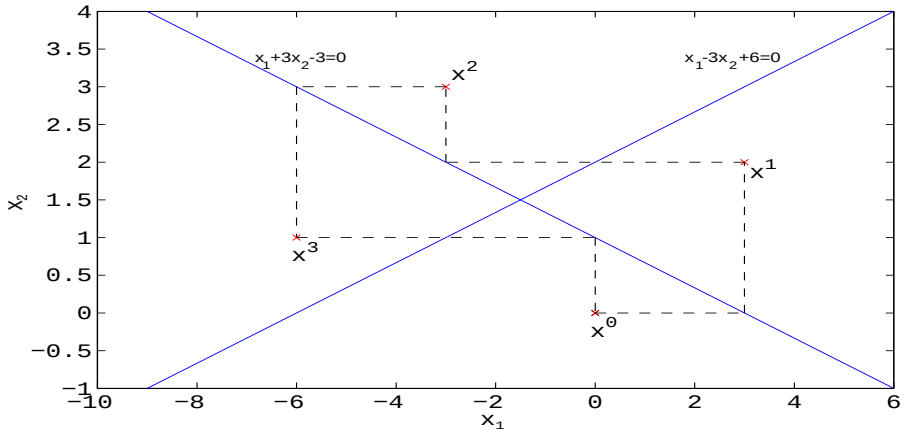
$$\begin{aligned} x_1 + x_2 &= 1 \\ qx_1 + x_2 &= 1. \end{aligned}$$

Odpovídající relaxační matice T_ω a vektor $\boldsymbol{g}_\omega$ jsou tvaru

$$T_\omega = \begin{pmatrix} 1-\omega & -\omega \\ -q\omega(1-\omega) & q\omega^2 + (1-\omega) \end{pmatrix}, \qquad \boldsymbol{g}_\omega = \begin{pmatrix} \omega \\ \omega(1-q\omega) \end{pmatrix}.$$



Obr. 5.5: Gaussova-Seidelova iterační metoda



Obr. 5.6: Jacobiova iterační metoda

Pro $\omega = 2$ je matice T_2 a vektor g_2 tvaru

$$T_2 = \begin{pmatrix} -1 & -2 \\ 2q & 4q-1 \end{pmatrix}, \qquad g_2 = \begin{pmatrix} 2 \\ 2(1-2q) \end{pmatrix}.$$

Charakteristický polynom

$$\varphi(\lambda) = \lambda^2 + \lambda(2-4q) + 1$$

má kořeny

$$\lambda_{1,2} = -1 + 2q \pm 2\sqrt{q(q-1)}.$$

Pro $q \in (0,1)$ jsou tyto kořeny komplexně sdružené a jejich absolutní hodnota je rovna jedné, tedy $\varrho(T_2) = 1$. Vlastní čísla λ_1, λ_2 mohou být také vyjádřena v goniometrickém tvaru

$$\lambda_{1,2} = \cos\varphi \pm i \sin\varphi$$

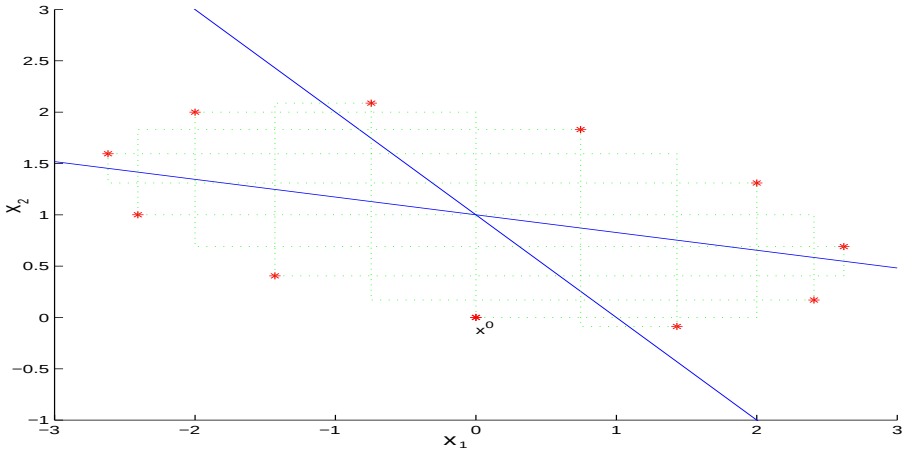
a lze ukázat ([24]), že

$$q = \frac{1}{2}(1 + \cos\varphi)\,.$$

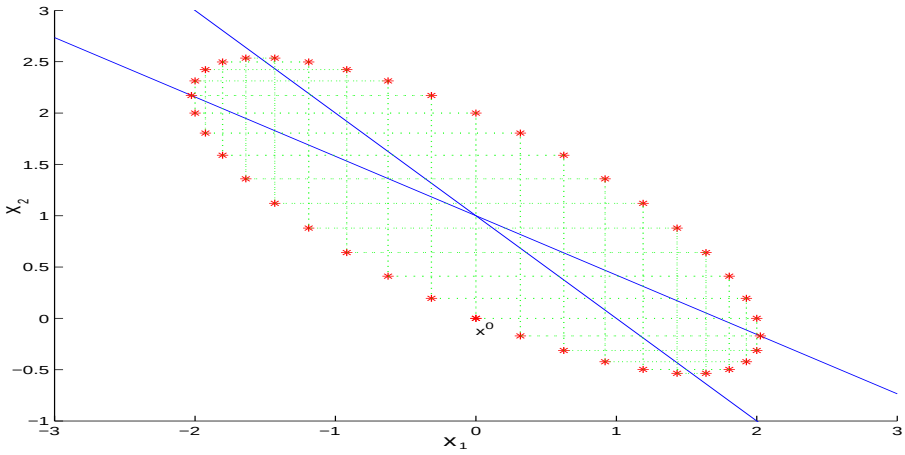
V práci [24] je dokázáno, že cyklus řádu $p > 2$ existuje tehdy a jen tehdy, když $\varphi = 2\pi l/p$, $0 < l < p/2$. Vztah $q = (1+\cos\varphi)/2$ umožňuje nalézt systém generující cyklus daného řádu p vhodnou volbou p, l.

Např.: pro $q = 0.5((1+\cos 2\pi\frac{4}{11}) \doteq 0.17257$ existuje cyklus řádu 11 (viz obr. 5.7)

pro $q = 0.5((1+\cos 2\pi\frac{9}{40}) \doteq 0.578217$ existuje cyklus řádu 40 (viz obr. 5.8).



Obr. 5.7: Cyklus řádu 11, počáteční aproximace $x^0 = (0,0)$.



Obr. 5.8: Cyklus řádu 40, počáteční aproximace $x^0 = (0, 0)$.

Všechny body cyklu řádu $p = l/s$ leží na elipse, jejíž střed je přesné řešení $\boldsymbol{x}^*$. Rovnice elipsy je tvaru

$$x_1^2 + \frac{x_2^2}{q} + 2x_1x_2 - \frac{2x_2}{q} - 2x_1 = (x_1^0)^2 + \frac{(x_2^0)^2}{q} + 2x_1^0x_2^0 - \frac{2x_2^0}{q} - 2x_1^0,$$

kde $\boldsymbol{x}^0 = (x_1^0, x_2^0)$ je počáteční aproximace.

Pro různou volbu počátečních aproximací dostáváme množinu soustředných elips.

Poznámka 6. Jestliže v iteračním procesu (5.4) nastává cyklus, pak $\varrho(T_\omega) = 1$. Ale opak neplatí. Je-li $\varrho(T_\omega) = 1$, nemusí v iteračním procesu nastat cyklus. Otázky cyklů iteračních metod jsou podrobně studovány v [24].

Cvičení ke kapitole 5

1. Je matice $H = \begin{pmatrix} 0,2 & 0,3 & -0,1 \\ 0,7 & -0,15 & 0,05 \\ 0,2 & 0,0 & 0,6 \end{pmatrix}$ konvergentní?

2. a) Jacobiovou, b) Gaussovou-Seidelovou iterační metodou řešte systémy

$$\begin{aligned} 10x_1 - 2x_2 - 2x_3 &= 6 \\ -x_1 + 10x_2 - 2x_3 &= 7 \\ -x_1 - x_2 + 10x_3 &= 8 \end{aligned} \qquad \begin{aligned} 2x_1 - x_2 + x_3 &= -1 \\ 3x_1 + 3x_2 + 9x_3 &= 0 \\ 3x_1 + 3x_2 + 5x_3 &= 4 \end{aligned} .$$

3. Ukažte, že pro systém

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2(1-\varepsilon)x_1 + x_2 + x_3 &= 2 \\ x_3 + x_4 &= -1 \\ -(1-\varepsilon)^2 x_1 + x_4 &= 5 \end{aligned}$$

$(0 < \varepsilon < 0{,}1)$ Jacobiova metoda konverguje a Gaussova-Seidelova metoda diverguje.

4. Ukažte, že pro systém

$$\begin{aligned} 2x_2 + 4x_3 &= 0 \\ x_1 - x_2 - x_3 &= 0{,}375 \\ x_1 - x_2 + 2x_3 &= 0 \end{aligned}$$

diverguje Jacobiova i Gaussova-Seidelova metoda.

5. Ukažte, že pro systém

$$\begin{aligned} 4x_1 + 3x_2 &= 24 \\ 3x_1 + 4x_2 - x_3 &= 30 \\ -x_2 + 4x_3 &= -24 \end{aligned}$$

Jacobiova iterační metoda konverguje. Zvolte počáteční aproximaci $\boldsymbol{x}^0 = (1,1,1)^T$ a vypočtěte $\boldsymbol{x}^1$ a $\boldsymbol{x}^2$.

6. Ukažte, že pro systém

$$\begin{aligned} 3x_1 + 2x_2 + 2x_3 &= 1 \\ 2x_1 + 3x_2 + 2x_3 &= 0 \\ 2x_1 + 2x_2 + 3x_3 &= -1 \end{aligned}$$

Jacobiova iterační metoda diverguje a Gaussova-Seidelova metoda konverguje.

7. Dokažte, že pro systém

$$\begin{aligned} 10x_1 - x_2 + 2x_3 - 3x_4 &= 0 \\ x_1 + 10x_2 - x_3 + 2x_4 &= 5 \\ 2x_1 + 3x_2 + 20x_3 - x_4 &= -10 \\ 3x_1 + 2x_2 + x_3 + 20x_4 &= 15 \end{aligned}$$

Jacobiova iterační metoda konverguje. Kolik iterací je třeba k nalezení řešení s chybou menší než 10^{-4}?
(Řešení: $k \geq 17$.)

8. Ukažte, že pro systém

$$\begin{aligned} 4x_1 + 3x_2 + 2x_3 &= -1 \\ x_1 + 5x_2 + 4x_3 &= 2 \\ 2x_1 - x_2 - 7x_3 &= 4 \end{aligned}$$

Jacobiova metoda konverguje. Zvolte počáteční iteraci $\boldsymbol{x}^0 = (1,1,1)^T$ a vypočtěte $\boldsymbol{x}^1$ a $\boldsymbol{x}^2$.

9. Ukažte, že pro systém

$$\begin{aligned} 3x_1 + x_2 + x_3 &= 2 \\ 2x_1 + 4x_2 + x_3 &= 4 \\ -x_1 - x_2 - 3x_3 &= -1 \end{aligned}$$

Gaussova-Seidelova metoda konverguje. Zvolte počáteční iteraci $x^0 = (0,0,0)^T$ a vypočtěte první dvě iterace.

10. Ukažte, že pro systém

$$\begin{aligned} 2x_1 - x_2 - x_3 &= 3 \\ -x_1 + 3x_2 - 2x_3 &= 1 \\ -3x_1 - x_2 + 4x_3 &= -1 \end{aligned}$$

Jacobiova iterační metoda diverguje.

11. Ukažte, že pro systém

$$\begin{aligned} 6x_1 + 3x_2 &= 10 \\ 3x_1 + 5x_2 - x_3 &= 5 \\ -x_2 + 4x_3 &= -6 \end{aligned}$$

Gaussova–Seidelova iterační metoda konverguje. Zvolte počáteční aproximaci $\boldsymbol{x}^0 = (1,1,1)^T$ a vypočtěte $\boldsymbol{x}^1$ a $\boldsymbol{x}^2$.

12. Systém

$$\begin{aligned} 4x_1 + 3x_2 &= 24 \\ 3x_1 + 4x_2 - x_3 &= 30 \\ - x_2 + 4x_3 &= -24 \end{aligned}$$

má přesné řešení $\boldsymbol{x}^* = (3,4,-5)^T$. Užijte Gaussovy-Seidelovy iterační metody a relaxační metody s parametrem $\omega = 1{,}25$. Porovnejte výsledky po provedení 7 iterací. Vypočtěte optimální hodnotu parametru ω.
(Řešení: $\varrho(T_G) = 0{,}625$, $\varrho(T_{\omega_{opt}}) \approx 0{,}24$, $\omega_{opt} \approx 1{,}24$.)

13. Systém

$$\begin{aligned} 10x_1 - \quad x_2 \qquad\quad &= 9 \\ -\ x_1 + 10x_2 - \ 2x_3 &= 7 \\ -\ 2x_2 + 10x_3 &= 6 \end{aligned}$$

(přesné řešení $\boldsymbol{x}^* = \left(\frac{473}{475}, \frac{455}{475}, \frac{376}{475}\right)^T$) řešte relaxační metodou s $\omega = 0{,}5$, $\omega = 1{,}1$ a vypočtěte optimální hodnotu parametru ω_{opt} a řešte systém relaxační metodou s tímto parametrem.

Kontrolní otázky ke kapitole 5

1. Může být v iteračním procesu $\boldsymbol{x}^{k+1} = T\boldsymbol{x}^k + \boldsymbol{g}$ iterační matice T singulární?
2. V příkladu 6 Jacobiova metoda pro systém

$$\begin{aligned} 3x_1 + 2x_2 + 2x_3 &= \ \ 1 \\ 2x_1 + 3x_2 + 2x_3 &= \ \ 0 \\ 2x_1 + 2x_2 + 3x_3 &= -1 \end{aligned}$$

obecně nekonverguje, neboť $\varrho(T_J) > 1$. Přesto pro počáteční aproximaci $x_0 = (0, 0, 0)^T$ iterační proces konverguje k řešení $(1, 0, -1)^T$. Proč?

3. Uvažujte systém

$$\begin{aligned} x_1 - x_2 &= 0 \\ x_1 + x_2 &= 0\,. \end{aligned}$$

Co znamená fakt, že $T_J^4 = E$?

4. Jestliže matice A není ryze řádkově nebo sloupcově diagonálně dominantní, může Jacobiova iterační metoda konvergovat?

Kapitola 6

Interpolace

V této kapitole budeme zkoumat problém aproximace funkcí. Tento problém spočívá většinou v nalezení aproximace funkce f pomocí vhodné kombinace funkcí z nějaké třídy funkcí. Uvažujme třídu funkcí jedné proměnné:

$$\psi(x; a_0, \ldots, a_n)\,,$$

kde $a_0, \ldots, a_n$ jsou parametry, jejichž hodnoty charakterizují jednotlivé funkce v této třídě. Ústředním problémem aproximace je kriterium pro volbu těchto parametrů. U *interpolační aproximace* požadujeme, aby parametry byly vybrány tak, že na množině navzájem různých bodů $\{x_i\}_{i=0}^n$ platí

$$\psi(x_i; a_0, \ldots, a_n) = f(x_i), \quad i = 0, \ldots, n\,.$$

V některých případech jsou předepsány i hodnoty derivací v některých bodech. V případě metody nejmenších čtverců se parametry $a_0, \ldots, a_n$ vyberou tak, aby veličina

$$\varrho(a_0, \ldots, a_n) = \sum_{j=0}^{N} (f(x_j) - \psi(x_j; a_0, \ldots, a_n))^2,\ n < N,$$

nabývala minimální hodnoty. Dalším typem aproximace je Čebyševova aproximace, kde hledáme takovou aproximaci, která minimalizuje maximální absolutní hodnotu rozdílu funkce f a aproximace ψ:

$$\min_{(a_0,\ldots,a_n)} \max_{x\in[a,b]} |f(x) - \psi(x; a_0, \ldots, a_n)|$$

V dalším se zaměříme pouze na lineární aproximaci, tj. budeme se zabývat funkcemi tvaru

$$\psi(x; a_0, \ldots, a_n) = a_0\psi_0(x) + \ldots + a_n\psi_n(x)\,,$$

kde funkce ψ_i, $i = 0, \ldots, n$ tvoří bázi lineárního prostoru dimenze $n + 1$. Do této třídy patří i klasická *polynomiální interpolace*

$$\psi(x; a_0, \ldots, a_n) = a_0x^n + a_1x^{n-1} + \ldots + a_n$$

a *trigonometrická interpolace*

$$\psi(x; a_1, \ldots, a_n) = a_0 + a_1 e^{\mathrm{i}x} + a_2 e^{2\mathrm{i}x} + \ldots + a_n e^{n\mathrm{i}x}, \qquad (\mathrm{i}^2 = -1)$$

Polynomiální interpolace se užívá především k aproximaci funkcí daných tabulkou a je také důležitým základem pro některé typy formulí numerického derivování a integrování. Do třídy lineárních interpolačních problémů také patří *splajnová interpolace*. Ve speciálním případě *kubických splajnů* se požaduje, aby funkce ψ byla dvakrát spojitě diferencovatelná pro $x \in [x_0, x_n]$ a byla totožná s kubickým polynomem na každém subintervalu $[x_i, x_{i+1}]$ daného dělení $x_0 < x_1 < \ldots < x_n$. Splajnové interpolaci je nyní věnována značná pozornost, protože poskytuje vhodný nástroj pro interpolaci empirických křivek a složitých matematických funkcí. Roste také její užití při přibližném řešení diferenciálních rovnic.

§ 6.1. Polynomiální interpolace

Nejjednodušší úlohu lineární interpolace lze formulovat takto: Jsou dány body x_i, $i = 0, 1, \ldots, n$, $x_i \neq x_k$ pro $i \neq k$ a hodnoty funkce f v těchto bodech: $f(x_i) = f_i$, $i = 0, 1, \ldots, n$. Je třeba najít algebraický polynom P_n stupně nejvýše n takový, že

$$P_n(x_i) = f_i, \qquad i = 0, 1, \ldots, n.$$

Úmluva. Body x_i, $i = 0, 1, \ldots, n$, $x_i \neq x_k$ pro $i \neq k$, budeme nazývat *uzly*, polynom P_n *interpolační polynom*. Jako dříve označme Π_n množinu všech reálných polynomů stupně nejvýše n tvaru

$$P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n.$$

Věta 6.1. *Pro* $(n+1)$ *daných dvojic čísel*

$$(x_i, f_i), \qquad i = 0, 1, \ldots, n, \quad x_i \neq x_k \text{ pro } i \neq k,$$

existuje právě jeden polynom $P_n \in \Pi_n$ *takový, že*

$$P_n(x_i) = f_i, \qquad i = 0, 1, \ldots, n. \tag{6.1}$$

Důkaz.
Jednoznačnost: Předpokládejme, že existují dva interpolační polynomy $P_n, Q_n \in \Pi_n$ splňující podmínky (6.1), tj.

$$P_n(x_i) = Q_n(x_i) = f_i, \qquad i = 0, 1, \ldots, n.$$

Položme $R_n(x) = P_n(x) - Q_n(x)$. Je zřejmé, že $R_n \in \Pi_n$ a dále $R_n(x_i) = 0$, $i = 0, 1, \ldots, n$, tzn. R_n má alespoň $n+1$ různých kořenů. To je ale spor s předpokladem, že R_n je polynom stupně nejvýše n. Odtud plyne, že polynomy P_n a Q_n musí být totožné.
Existence: Existenci dokážeme tak, že příslušný polynom sestrojíme. Nejdříve sestrojíme polynomy l_i, $i = 0, 1, \ldots, n$ s těmito vlastnostmi:

(a) l_i je polynom stupně n,

(b) $l_i(x_j) = \begin{cases} 0 & \text{pro } i \neq j \\ 1 & \text{pro } i = j. \end{cases}$

Je zřejmé, že

$$l_i(x) = A_i(x - x_0)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n).$$

Konstantu A_i určíme tak, aby byla splněna podmínka $l_i(x_i) = 1$, tedy

$$A_i = \frac{1}{(x_i - x_0)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)}.$$

a odtud

$$l_i(x) = \frac{(x - x_0)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n)}{(x_i - x_0)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)}. \tag{6.2}$$

Definujme nyní polynom P_n vztahem:

$$P_n(x) = l_0(x)f_0 + l_1(x)f_1 + \dots + l_n(x)f_n = \sum_{i=0}^{n} l_i(x)f_i. \tag{6.3}$$

Snadno se ověří, že tento polynom splňuje interpolační podmínky (6.1), a protože je lineární kombinací polynomů stupně n, je polynomem stupně nejvýše n. □

Interpolační polynom tvaru (6.3) nazýváme *Lagrangeovým* interpolačním polynomem nebo přesněji Lagrangeovým tvarem interpolačního polynomu.

Úmluva. Polynomy l_i, $i = 0, 1, \dots, n$, definované vztahem (6.2) budeme nazývat *fundamentální polynomy*.

Z jednoznačnosti interpolačního polynomu rovněž plyne, že interpolační polynom stupně nejvýše n pro polynom Q_n stupně n je tentýž polynom, tj. $P_n(x) \equiv Q_n(x)$.

Položme

$$\omega_{n+1}(x) = (x - x_0)\dots(x - x_n) = \prod_{i=0}^{n}(x - x_i).$$

Je zřejmé, že

$$\omega'_{n+1}(x_i) = (x_i - x_0)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n) = \prod_{\substack{k=0 \\ k \neq i}}^{n} (x_i - x_k).$$

Použitím těchto vztahů lze fundamentální polynomy zapsat ve tvaru

$$l_i(x) = \frac{\omega_{n+1}(x)}{(x - x_i)\,\omega'_{n+1}(x_i)}, \qquad i = 0, 1, \dots, n. \tag{6.4}$$

Příklad 6.1. Sestrojte Lagrangeův interpolační polynom, je-li dáno:

x_i	0	1	2	5
f_i	2	3	12	147

Řešení. V tomto případě je $n = 3$, hledáme tedy polynom $P_3 \in \Pi_3$.

$$\begin{aligned} P_3(x) &= 2\frac{(x-1)(x-2)(x-5)}{(-1)(-2)(-5)} + 3\frac{(x-0)(x-2)(x-5)}{(1-0)(1-2)(1-5)} + \\ &\quad + 12\frac{(x-0)(x-1)(x-5)}{(2-0)(2-1)(2-5)} + 147\frac{(x-0)(x-1)(x-2)}{(5-0)(5-1)(5-2)} = \\ &= x^3 + x^2 - x + 2. \end{aligned}$$

Příklad 6.2. Sestrojte interpolační polynom pro funkci $f(x) = x + \sin x$ v uzlech $x_0 = 9$, $x_1 = 3$, $x_2 = 4{,}5$, $x_3 = 10$, $x_4 = 5{,}5$, $x_5 = 12{,}5$.
Řešení. Jelikož $n = 5$, hledáme polynom $P_5 \in \Pi_5$. Na obr. 6.1 jsou znázorněny fundamentální polynomy l_i, $i = 0, 1, \ldots 5$, na obr. 6.2 je znázorněn čárkovaně polynom P_5 a plnou čarou je znázorněn graf dané funkce.

Fundamentální polynomy l_i, $i = 0, \ldots, n$ mají zajímavou vlastnost, splňují totiž identitu

$$\sum_{i=0}^{n} l_i(x) = 1 \quad \forall x \in \mathbb{R}\,.$$

Toto tvrzení plyne z faktu, že interpolační polynom $P_n \in \Pi_n$ pro funkci f, která je polynomem stupně nejvýše n, platí $P_n(x) = f(x)$, $\forall x \in \mathbb{R}$. Tedy i pro funkci $f(x) \equiv 1$, která je polynomem nultého stupně, tvrzení platí. Uvedený vztah dále vyjadřuje skutečnost, že polynomy l_i, $i = 0, \ldots, n$ jsou jisté „váhy“ přiřazené hodnotám f_i, $i = 0, \ldots, n$.

Další zajímavou vlastnost interpolačního polynomu dostaneme pomocí následujícího výpočtu:
Platí

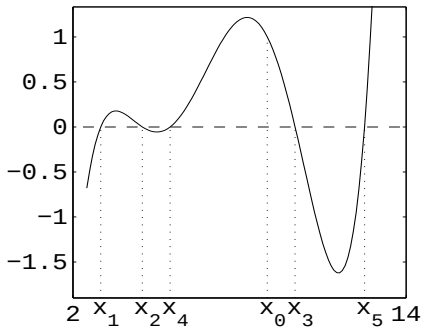
$$P_n(x) = \sum_{i=0}^{n} \frac{\omega_{n+1}(x)}{(x - x_i)\,\omega'_{n+1}(x_i)}\, f(x_i)\,.$$

Označme $w_i = (\omega'_{n+1}(x_i))^{-1}$, tedy w_i závisí pouze na x_i a platí

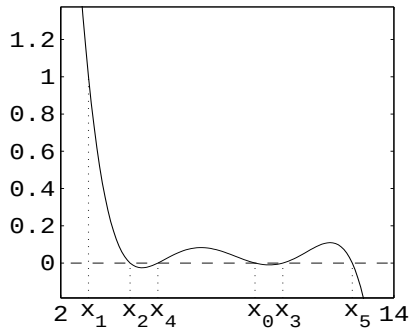
$$P_n(x) = \omega_{n+1}(x) \sum_{i=0}^{n} \frac{f(x_i)\, w_i}{(x - x_i)}\,.$$

Dále víme, že

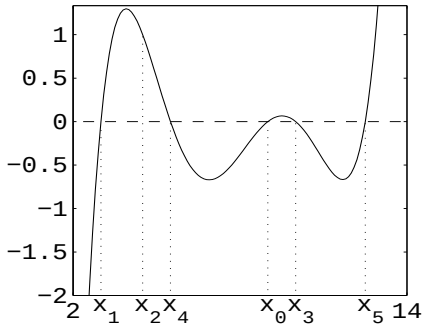
$$\sum_{i=0}^{n} \frac{\omega_{n+1}(x)}{(x - x_i)\,\omega'_{n+1}(x_i)} = 1\,,$$



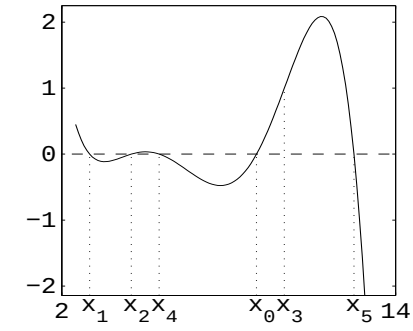
Fundamentální polynom $l_0(x)$



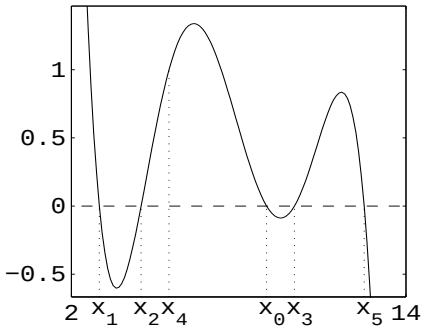
Fundamentální polynom $l_1(x)$



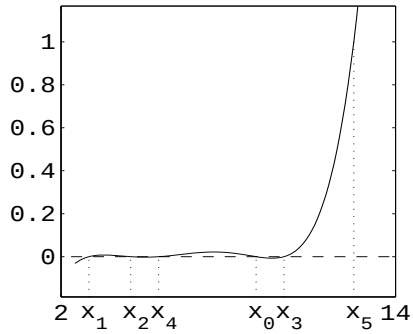
Fundamentální polynom $l_2(x)$



Fundamentální polynom $l_3(x)$

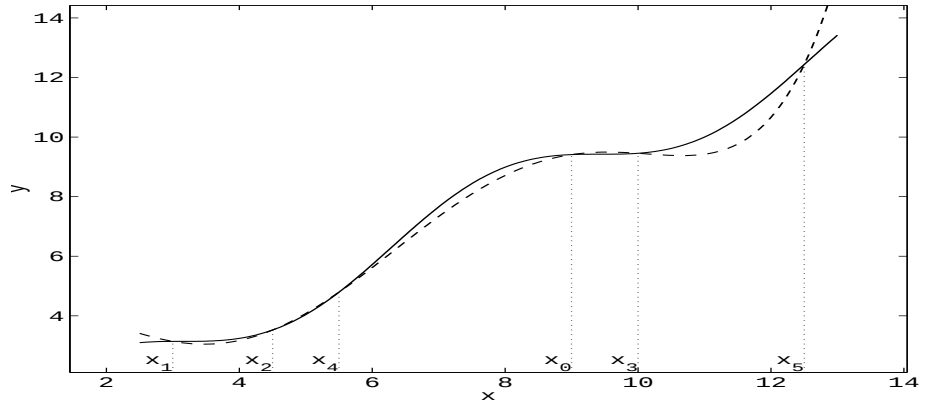


Fundamentální polynom $l_4(x)$



Fundamentální polynom $l_5(x)$

Obr. 6.1: Fundamentální polynomy l_i



Obr. 6.2: Lagrangeův interpolační polynom pro $f(x) = x + \sin x$ v bodech (x_i, f_i), $i = 0, \ldots, 5$

tj.

$$\omega_{n+1}(x) \sum_{i=0}^{n} \frac{w_i}{(x - x_i)} = 1\,,$$

neboli

$$\omega_{n+1}(x) = \frac{1}{\sum\limits_{i=0}^{n} \frac{w_i}{(x-x_i)}}\,.$$

Odtud

$$P_n(x) = \frac{\sum\limits_{i=0}^{n} \frac{f(x_i)\, w_i}{(x-x_i)}}{\sum\limits_{i=0}^{n} \frac{w_i}{(x-x_i)}} = \sum_{i=0}^{n} \frac{\frac{w_i}{(x-x_i)}}{\sum\limits_{j=0}^{n} \frac{w_j}{(x-x_j)}} f(x_i)\,.$$

Hodnoty $\frac{w_i}{(x-x_i)} \Big/ \sum\limits_{j=0}^{n} \frac{w_j}{(x-x_j)}$ tvoří tzv. *barycentrické* souřadnice bodu $(x, P_n(x))$ v rovině vzhledem k bodům $(x_i, f(x_i))$, $i = 0, \ldots, n$. Tato formule je navíc z výpočetního hlediska velmi efektivní – při určených hodnotách w_i je výpočet hodnoty $P_n(x)$ lineární vzhledem k n. Problémy mohou ale nastat pro x blízké některému uzlu x_i, kde při dělení výrazem $x - x_i$, který je blízký nule, dochází k velké relativní chybě.

Poznámka 1. Danou úlohu interpolace lze rovněž řešit metodou neurčitých koeficientů. Tato metoda spočívá v následujícím. Podmínky (6.1) zapíšeme ve tvaru systému lineárních rovnic pro neznámé koeficienty a_i, $i = 0, 1, 2, \ldots, n$,

$$a_0 x_i^n + a_1 x_i^{n-1} + \ldots + a_n = 0, \qquad i = 0, 1, \ldots, n.$$

Jestliže body x_i, $i = 0, \ldots, n$, jsou navzájem různé, má tento systém právě jedno řešení. Ale tento postup je příliš těžkopádný a nevhodný pro větší počet uzlů.

Nechť nyní P_{n-1} je Lagrangeův interpolační polynom v uzlech $x_0, \ldots, x_{n-1}$. Rozdíl $f(x) - P_{n-1}(x)$ vydělme součinem $(x - x_0) \ldots (x - x_{n-1})$. Tento podíl v bodě $x = x_n$ vyjádříme ve tvaru

$$\begin{aligned}\frac{f(x_n) - P_{n-1}(x_n)}{\prod_{j=0}^{n-1}(x_n - x_j)} &= \frac{f(x_n) - P_{n-1}(x_n)}{\omega_n(x_n)} = \\ &= \frac{1}{\omega_n(x_n)} \left(f(x_n) - \sum_{j=0}^{n-1} f(x_j) \frac{\omega_n(x_n)}{\omega_n'(x_n)(x_n - x_j)} \right) = \\ &= \sum_{j=0}^{n} \frac{f(x_j)}{\prod\limits_{\substack{i=0 \\ i \neq j}}^{n} (x_j - x_i)} = f[x_0, \ldots, x_n]. \end{aligned} \tag{6.5}$$

Definice 6.1. Výraz $f[x_0, \ldots, x_n]$ ve vztahu (6.5) nazýváme *poměrnou diferencí řádu n* funkce f v bodech $x_0, \ldots, x_n$.

Položme $f(x_0) = f[x_0]$. Dále

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

Poznamenejme, že *poměrná diference je symetrickou funkcí svých argumentů*, tedy hodnota poměrné diference nezávisí na pořadí uzlů x_i.

Lagrangeův interpolační polynom je po teoretické stránce velmi důležitý. Je základem pro odvození metod numerického derivování a integrování. Ale pro praktické výpočty, zejména pro velký počet uzlů nebo při změně počtu uzlů (kdy je třeba přepočítat všechny polynomy l_i), je výhodnější použít některé z formulí, které nyní uvedeme.

Věta 6.2. *Interpolační polynom $P_n \in \Pi_n$ pro body (x_i, f_i), $i = 0, 1, \ldots, n$, může být zapsán ve tvaru*

$$P_n(x) = f_0 + (x - x_0) f[x_0, x_1] + \ldots + (x - x_0) \ldots (x - x_{n-1}) f[x_0, \ldots, x_n]. \tag{6.6}$$

Důkaz. Budeme postupovat tak, že vhodným způsobem vyjádříme Lagrangeův interpolační polynom a další úpravou dostaneme vyjádření (6.6). Nechť $P_n \in \Pi_n$ je Lagrangeův interpolační polynom pro dané body (x_i, f_i), $i = 0, 1, \ldots, n$. Zapišme tento polynom ve tvaru

$$\begin{aligned} P_n(x) = &P_0(x) + [P_1(x) - P_0(x)] + \ldots + [P_j(x) - P_{j-1}(x)] + \ldots \\ &\ldots + [P_n(x) - P_{n-1}(x)], \end{aligned} \tag{6.7}$$

kde $P_j \in \Pi_j$ je Lagrangeův interpolační polynom pro body (x_i, f_i), $i = 0, 1, 2, \ldots, j$. Počítejme rozdíly $P_j(x) - P_{j-1}(x)$:

$$P_j(x) - P_{j-1}(x) = \sum_{i=0}^{j} \frac{\omega_{j+1}(x)}{(x - x_i)\,\omega'_{j+1}(x_i)} f_i - \sum_{i=0}^{j-1} \frac{\omega_j(x)}{(x - x_i)\,\omega'_j(x_i)} f_i =$$

$$= \frac{\omega_{j+1}(x)}{(x - x_j)\,\omega'_{j+1}(x_j)} f_j + \sum_{i=0}^{j-1} f_i \left[\frac{\omega_{j+1}(x)}{(x - x_i)\,\omega'_{j+1}(x_i)} - \frac{\omega_j(x)}{(x - x_i)\,\omega'_j(x_i)} \right].$$

Pro ω_{j+1} a ω_j platí

$$\omega_{j+1}(x) = \prod_{i=0}^{j} (x - x_i), \quad \omega_j(x) = \prod_{i=0}^{j-1} (x - x_i) \quad \Rightarrow \quad \omega_{j+1}(x) = (x - x_j)\,\omega_j(x).$$

Dále pro derivace funkcí ω_j a ω_{j+1} máme

$$\omega'_{j+1}(x) = \omega_j(x) + (x - x_j)\,\omega'_j(x) \quad \Rightarrow \quad \omega'_{j+1}(x_i) = \begin{cases} (x_i - x_j)\,\omega'_j(x_i) & \text{pro } i \neq j \\ \omega_j(x_j) & \text{pro } i = j. \end{cases}$$

Těchto vztahů nyní užijeme pro výpočet rozdílů $P_j(x) - P_{j-1}(x)$:

$$\begin{aligned} P_j(x) - P_{j-1}(x) &= \frac{\omega_{j+1}(x)}{(x - x_j)\,\omega'_{j+1}(x_j)} f_j + \sum_{i=0}^{j-1} \frac{\omega_j(x)(x - x_j - x_i + x_j)}{(x - x_i)\,\omega'_{j+1}(x_i)} f_i = \\ &= \frac{\omega_j(x)}{\omega'_{j+1}(x_j)} f_j + \omega_j(x) \sum_{i=0}^{j-1} \frac{f_i}{\omega'_{j+1}(x_i)} = \omega_j(x) \sum_{i=0}^{j} \frac{f_i}{\omega'_{j+1}(x_i)} = \\ &= \omega_j(x) f[x_0, \ldots, x_j] \end{aligned}$$

Každý rozdíl $P_j(x) - P_{j-1}(x)$ pro $j = 0, 1, \ldots, n$ můžeme tedy vyjádřit ve tvaru

$$P_j(x) - P_{j-1}(x) = (x - x_0) \ldots (x - x_{j-1}) f[x_0, \ldots, x_j].$$

Odtud a z (6.7) nyní plyne, že interpolační polynom může být zapsán ve tvaru (6.6). □

Interpolační polynom (6.6) se nazývá *Newtonův interpolační polynom*.

Ještě jednou připomínáme, že pro dané body (x_i, f_i), $i = 0, 1, \ldots, n$, $x_i \neq x_k$ pro $i \neq k$, je interpolační polynom určen *jednoznačně*, tzn., že Newtonův interpolační polynom je totožný s Lagrangeovým interpolačním polynomem, liší se pouze formou zápisu.

Důsledek 1. *Nechť $x_i \in [a, b]$, $i = 0, \ldots, n$, $x_i \neq x_k$ pro $i \neq k$. Nechť $f \in C^n[a, b]$. Pak existuje bod $\theta \in (a, b)$ takový, že*

$$f[x_0, \ldots, x_n] = \frac{1}{n!} f^{(n)}(\theta). \tag{6.8}$$

Důkaz. Funkce $\Phi(x) = f(x) - P_n(x)$ má alespoň $(n+1)$ nulových bodů v $[a,b]$: $x_0, \dots, x_n$. Podle Rolleovy věty má funkce $\Phi^{(n)}(x) = f^{(n)}(x) - P_n^{(n)}(x)$ alespoň jeden nulový bod $\theta \in (a,b)$:

$$f^{(n)}(\theta) - P_n^{(n)}(\theta) = 0,$$

a

$$f^{(n)}(\theta) = P_n^{(n)}(\theta).$$

Na druhé straně, ze vztahu (6.6) plyne

$$P_n^{(n)}(\theta) = n! f[x_0, \dots, x_n]$$

a odtud

$$\frac{1}{n!} f^{(n)}(\theta) = f[x_0, \dots, x_n].$$

□

Poznámka 2. Lze ukázat ([23]), že

$$\lim_{\substack{x_i \to x_0 \\ i=1,\dots,n}} f[x_0, \dots, x_n] = \frac{1}{n!} f^{(n)}(x_0)\,, \tag{6.9}$$

neboli

$$f[\underbrace{x_0, \dots x_0}_{(n+1)\text{krát}}] = \frac{1}{n!} f^{(n)}(x_0)\,. \tag{6.10}$$

Důsledek 2. *Nechť Q_{n-1} je polynom stupně nejvýše $n-1$. Pak*

$$Q_{n-1}[x_0, \dots, x_n] = 0. \tag{6.11}$$

Je-li $Q_n(x) = x^n$, pak

$$Q_n[x_0, \dots, x_n] = 1. \tag{6.12}$$

Důkaz. (6.11) plyne ihned ze skutečnosti, že interpolační polynom pro Q_{n-1} je tentýž polynom a tedy poslední člen ve vyjádření (6.6) musí být roven nule. Vztah (6.12) plyne z následujícího: Nechť P_{n-1} je interpolační polynom pro funkci $Q_n(x) = x^n$ v uzlech $x_0, \dots, x_{n-1}$. Rozdíl $x^n - P_{n-1}(x)$ je tedy polynom stupně n s kořeny v bodech $x_0, \dots, x_{n-1}$, což znamená, že tento rozdíl lze vyjádřit ve tvaru

$$x^n - P_{n-1}(x) = (x - x_0) \dots (x - x_{n-1}),$$

Dále ze vztahu (6.5) pro poměrnou diferenci plyne pro $x = x_n$

$$1 = \frac{x_n^n - P_{n-1}(x_n)}{\prod_{i=0}^{n-1} (x_n - x_i)} = Q_n[x_0, \dots, x_n].$$

□

Lemma. *Platí identita*

$$(x_0 - x_n) f[x_0, \dots, x_n] = f[x_0, \dots, x_{n-1}] - f[x_1, \dots, x_n]. \tag{6.13}$$

Důkaz. Nechť $P_n \in \Pi_n$ je interpolační polynom splňující podmínky (6.1). Ze vztahu (6.6) plyne

$$P_n^{(n-1)}(x) = f[x_0,\ldots,x_{n-1}](n-1)! + \\ + f[x_0,\ldots,x_n](n!x - (n-1)!(x_0+\ldots+x_{n-1})). \tag{6.14}$$

Vyměníme-li ve vztahu (6.6) body x_0 a x_n, dostaneme

$$\begin{aligned} P_n^{(n-1)}(x) &= f[x_n,x_1,\ldots,x_{n-1}](n-1)! + \\ &\quad + f[x_n,x_1,\ldots,x_{n-1},x_0](n!x - (n-1)!(x_n+x_1+\ldots+x_{n-1})) = \\ &= f[x_1,\ldots,x_n](n-1)! + \\ &\quad + f[x_0,\ldots,x_n](n!x - (n-1)!(x_1+\ldots x_n)). \end{aligned} \tag{6.15}$$

Odečtením vztahů (6.14) a (6.15) dostaneme požadovanou identitu (6.13). □

Poznámka 3. Uvedené lemma znamená, že *poměrnou diferenci n-tého řádu lze rekurentně vyjádřit pomocí diferencí řádu* $n-1$:

$$f[x_0,\ldots,x_n] = \frac{f[x_1,\ldots,x_n] - f[x_0,\ldots,x_{n-1}]}{x_n - x_0} \tag{6.16}$$

Na základě tohoto rekurentního vztahu lze sestavit následující tabulku poměrných diferencí:

x_i	f_i	$f[x_i,x_{i+1}]$	$f[x_i,x_{i+1},x_{i+2}]$	$\ldots$	
x_0	$\underline{f_0}$	$\underline{f[x_0,x_1]}$	$\underline{f[x_0,x_1,x_2]}$		
x_1	f_1	$f[x_1,x_2]$	$\vdots$	$\cdots$	$\underline{f[x_0,\ldots,x_n]}$
x_2	f_2	$\vdots$	$f[x_{n-2},x_{n-1},x_n]$		
$\vdots$	$\vdots$	$f[x_{n-1},x_n]$			
x_n	f_n				

Je jasné, že pro konstrukci Newtonova interpolačního polynomu jsme užili hodnot označených $\underline{\quad}$.

Příklad 6.3. Pro hodnoty uvedené v příkladu 6.1 sestrojte Newtonův interpolační polynom.
Řešení. Sestavíme podle vztahu (6.16) tabulku poměrných diferencí:

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	$\underline{2}$	$\underline{1}$		
1	3	9	$\underline{4}$	
2	12	45	9	$\underline{1}$
5	147			

$$P_3(x) = 2 + x + 4(x-1)x + x(x-1)(x-2) = x^3 + x^2 - x + 2.$$

Pro výpočet Newtonova polynomu jsme užili *hodnot ležících na „diagonále“*.

Připomeňme ještě, že z definice poměrné diference (vztah (6.5)) plyne, že *poměrná diference závisí lineárně na funkci f: pro libovolná reálná čísla a, b a funkce f, g definované v bodech x_i, $i = 0, 1, \ldots, x_n$, platí*:

$$(af + bg)[x_0, \ldots, x_n] = a(f[x_0, \ldots, x_n]) + b(g[x_0, \ldots, x_n])$$

§ 6.2. Chyba interpolace

Zabývejme se nyní otázkou, s jakou přesností bude interpolační polynom $P_n \in \Pi_n$ aproximovat danou funkci v bodech různých od bodů x_i, $i = 0, 1, \ldots, n$. Jaký bude rozdíl $E(\bar{x}) = f(\bar{x}) - P_n(\bar{x})$ pro $\bar{x} \neq x_i$, $i = 0, 1, \ldots, n$? Odpověď dává následující věta:

Věta 6.3. *Nechť $f \in C^{(n+1)}[a, b]$ a nechť uzly $x_i \in [a, b]$, $i = 0, 1, \ldots, n$, $x_i \neq x_k$ pro $i \neq k$. Nechť dále $P_n \in \Pi_n$ je interpolační polynom splňující podmínky* (6.1). *Pak ke každému bodu $\bar{x} \in [a, b]$ existuje bod $\xi \in (a, b)$ tak, že platí:*

$$f(\bar{x}) - P_n(\bar{x}) = \frac{\omega_{n+1}(\bar{x})}{(n+1)!} f^{(n+1)}(\xi), \qquad \xi = \xi(\bar{x}). \tag{6.17}$$

Důkaz. Sestrojme Newtonův tvar interpolačního polynomu podle vztahu (6.6) pro uzly $x_0, \ldots, x_n$, $\bar{x}$. Je tedy třeba najít interpolační polynom $P_{n+1} \in \Pi_{n+1}$. Tento polynom je podle (6.6) tvaru

$$P_{n+1}(x) = f_0 + (x - x_0) f[x_0, x_1] + \ldots + (x - x_0) \ldots (x - x_n) f[x_0, \ldots, x_n, \bar{x}].$$

Jelikož P_{n+1} je interpolačním polynomem i v bodě $\bar{x}$, je $f(\bar{x}) = P_{n+1}(\bar{x})$. Ale na druhé straně

$$P_{n+1}(\bar{x}) = P_n(\bar{x}) + (\bar{x} - x_0) \ldots (\bar{x} - x_n) f[x_0, \ldots, x_n, \bar{x}],$$

neboli

$$f(\bar{x}) - P_n(\bar{x}) = (\bar{x} - x_0) \ldots (\bar{x} - x_n) f[x_0, \ldots, x_n, \bar{x}].$$

Zde

$$\omega_{n+1}(\bar{x}) = \prod_{i=0}^{n} (\bar{x} - x_i).$$

Nyní podle důsledku 1 je

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

a odtud plyne tvrzení. Vztah (6.17) jsme dokázali použitím Newtonova interpolačního polynomu. Ale z jednoznačnosti interpolačního polynomu plyne, že vztah platí pro polynom vyjádřený v libovolném tvaru. □

Poznámka 4. Z důsledku 1 je jasné, že bod ξ závisí na $\bar{x}$. Této skutečnosti si musíme být vědomi při dalších úvahách a operacích týkajících se chyby interpolace. Rozdíl $E(\bar{x}) = f(\bar{x}) - P_n(\bar{x})$ nazýváme *chybou interpolace v bodě $\bar{x}$*.

Jestliže $|f^{(n+1)}(x)| \le M_{n+1}$, $\forall x \in [a,b]$, lze chybu interpolace ohraničit shora takto

$$|E(\bar{x})| \le \frac{M_{n+1}}{(n+1)!} |\omega_{n+1}(\bar{x})| .$$

Tento odhad závisí na vlastnostech interpolované funkce a na volbě uzlů x_i. Vzniká tedy otázka, jak volit uzly x_i, aby maximální absolutní hodnota ω_{n+1} byla na daném intervalu co nejmenší. Toho lze dosáhnout tak, že za uzly x_i zvolíme kořeny některých speciálních polynomů. O tomto přístupu nyní stručně pojednáme.

Úmluva. Třídu všech normovaných polynomů stupně m, tj. polynomů tvaru

$$x^m + a_{m-1}x^{m-1} + \ldots + a_0$$

označíme $\bar{\Pi}_m$.

Definice 6.2. Řekneme, že polynom $Q_m \in \bar{\Pi}_m$ má ze všech polynomů třídy $\bar{\Pi}_m$ nejmenší odchylku od nuly na intervalu $[-1,1]$, jestliže platí

$$\max_{-1\le x\le 1} |Q_m(x)| < \max_{-1\le x\le 1} |S_m(x)|$$

pro všechny polynomy $S_m \in \bar{\Pi}_m$.

Věta 6.4. *$T_m(x) = \cos(m \arccos x)$, $x \in [-1,1]$ je polynom stupně m s koeficientem 2^{m-1} u x^m, $m \ge 1$.*

Důkaz. Je $T_0(x) \equiv 1$, $T_1(x) = x$. Dále použijeme vztahu

$$\cos(m+1)\alpha + \cos(m-1)\alpha = 2\cos\alpha\cos m\alpha.$$

Položíme-li $\alpha = \arccos x$, dostaneme

$$T_{m+1}(x) + T_{m-1}(x) = 2xT_m(x),$$

tj.

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x). \tag{6.18}$$

Nechť nyní podle indukčního předpokladu je T_m polynom stupně m s koeficientem 2^{m-1} u x^m. Pak z rekurentního vztahu (6.18) ihned plyne, že T_{m+1} je polynom stupně $m+1$ s koeficientem 2^m u x^{m+1}. □

Uvedeme nyní některé důležité vlastnosti polynomů T_m. Z rekurentního vztahu (6.18) plyne

$$T_2(x) = 2x^2 - 1, \qquad T_3(x) = 4x^3 - 3x, \qquad T_4(x) = 8x^4 - 8x^2 + 1, \qquad \text{atd.}$$

Polynom T_m má m kořenů, které jsou reálné, různé a všechny leží v intervalu $(-1,1)$, neboť pro ně platí

$$\begin{aligned} \cos(m\arccos x_k) &= 0 \\ m\arccos x_k &= \frac{2k+1}{2}\pi, \qquad k = 0,1,\ldots,m-1 \\ x_k &= \cos\frac{2k+1}{2m}\pi, \qquad k = 0,1,\ldots,m-1 \end{aligned} \tag{6.19}$$

Dále je

$$\max_{-1\le x\le 1} |T_m(x)| = 1.$$

Této maximální hodnoty nabývá T_m se střídavými znaménky v $(m+1)$ různých bodech intervalu $[-1,1]$:

$$|\cos(m \arccos x_k)| = 1$$
$$x_k = \cos\frac{k\pi}{m}, \qquad k = 0,1,\dots,m \tag{6.20}$$

Věta 6.5. *Polynom $\bar{T}_m(x) = 2^{1-m}T_m(x)$, $m \ge 1$, má na intervalu $[-1,1]$ nejmenší odchylku od nuly ze všech polynomů třídy $\bar{\Pi}_m$.*

Důkaz. Předpokládejme, že existuje polynom $Q_m \in \bar{\Pi}_m$, který má na intervalu $[-1,1]$ menší absolutní hodnotu než polynom $\bar{T}_m$. Uvažujme polynom

$$R_{m-1}(x) = \bar{T}_m(x) - Q_m(x).$$

Je zřejmě $R_{m-1} \in \Pi_{m-1}$ a v bodech $x_k = \cos(k\pi/m)$, $k = 0,1,\dots,m$ platí

$$\operatorname{sign}\left(\bar{T}_m(x_k) - Q_m(x_k)\right) = \operatorname{sign}\left(2^{1-m}(-1)^k - Q_m(x_k)\right) = (-1)^k$$

neboť podle předpokladu $|\bar{T}_m(x_k)| = 2^{1-m}$, $|Q_m(x_k)| < 2^{1-m}$. Polynom R_{m-1} mění tedy znaménko mezi body x_k, x_{k+1}, $k = 0,1,\dots m-1$. Odtud plyne, že tento polynom stupně $m-1$ má m různých kořenů. Dospěli jsme ke sporu. Předpokládejme nyní, že existuje polynom Q_m s maximální absolutní hodnotou rovnou maximální absolutní hodnotě polynomu $\bar{T}_m$. Není-li aspoň v jednom bodě, ve kterém nabývá polynom $\bar{T}_m$ extrému, polynom Q_m roven polynomu $\bar{T}_m$, dostaneme spor jako výše. Je-li však v takovém bodě $Q_m(x) = \bar{T}_m(x)$, má polynom R_{m-1} v tomto bodě dvojnásobný kořen a určíme-li počet kořenů polynomu R_{m-1} jako výše, dojdeme opět ke sporu, který dokončuje důkaz věty (viz [19]). □

Definice 6.3. Polynomy T_m se nazývají *Čebyševovy polynomy*.

Jestliže se při interpolaci omezíme na interval $[-1,1]$ a za uzly interpolace zvolíme kořeny Čebyševova polynomu,$\omega_{n+1}(x) = 2^{-n}T_{n+1}(x)$, můžeme najít odhad chyby ve tvaru

$$|f(\bar{x}) - P_n(\bar{x})| \le \frac{M_{n+1}}{(n+1)!2^n}, \tag{6.21}$$

neboť

$$\max_{x\in[-1,1]} |\omega_{n+1}(x)| = 2^{-n}.$$

Chyba je v tomto případě nejmenší možná. Při interpolaci na libovolném intervalu $[a,b]$ užijeme lineární transformaci

$$x = \frac{1}{2}((b-a)z + b + a), \qquad z \in [-1,1].$$

Při této transformaci se kořeny z_k, $k = 0, 1, \ldots, n$, polynomu $\bar{T}_{n+1}$ transformují na kořeny

$$x_k = \frac{1}{2}((b-a)z_k + b + a).$$

Odhad chyby interpolace nyní bude

$$|f(\bar{x}) - P_n(\bar{x})| \leq \frac{M_{n+1}}{(n+1)!}\frac{(b-a)^{n+1}}{2^{2n+1}}.$$

K dalším otázkám týkajících se chyby interpolace se vrátíme v další části této kapitoly.

§ 6.3. Interpolace na ekvidistantních uzlech

Nyní předpokládejme, že body x_i, $i = 0, 1, \ldots, n$, jsou *ekvidistantní*, tj. existuje reálné číslo $h \neq 0$ takové, že

$$x_i = x_0 + ih, \qquad i = 0, \ldots, n.$$

Číslo h obvykle nazýváme *krok*. Vypočteme pro Lagrangeův interpolační polynom $P_{n-1} \in \Pi_{n-1}$, který je interpolačním polynomem v uzlech $x_0, \ldots, x_{n-1}$, fundamentální polynomy l_i:

$$l_i(x) = \frac{\omega_n(x)}{(x - x_i)\,\omega_n'(x_i)}$$

Počítejme hodnoty těchto polynomů v bodě x_n:

$$\begin{aligned} l_i(x_n) &= \frac{\omega_n(x_n)}{(x_n - x_i)\,\omega_n'(x_i)} = \\ &= \frac{(x_n - x_0)\ldots(x_n - x_{n-1})}{(x_n - x_i)(x_i - x_0)\ldots(x_i - x_{i-1})(x_i - x_{i+1})\ldots(x_i - x_{n-1})} \end{aligned}$$

Nyní je $x_n - x_i = x_0 + nh - (x_0 + ih) = (n-i)h$. Dosazením do předchozího vztahu dostaneme:

$$\begin{aligned} l_i(x_n) &= \frac{n(n-1)\ldots(n-(n-1))}{(n-i)(i)\ldots(i-(i-1))(i-(i+1))\ldots(i-(n-1))} = \\ &= -\frac{n!}{i!(n-i)!}(-1)^{n-i} \end{aligned}$$

Odtud

$$l_i(x_n) = -(-1)^{n-i}\binom{n}{i}.$$

Počítejme nyní rozdíl

$$f(x_n) - P_{n-1}(x_n) = f(x_n) + \sum_{i=0}^{n-1}(-1)^{n-i}\binom{n}{i}f(x_i).$$

Hodnotu $f(x_n)$ lze zahrnout do součtu s koeficientem

$$(-1)^{n-n}\binom{n}{n} = 1$$

a výsledkem je

$$f(x_n) - P_{n-1}(x_n) = \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i} f_i. \tag{6.22}$$

Definice 6.4. Výraz

$$\Delta^n f_j = \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i} f_{i+j} \tag{6.23}$$

se nazývá n-tá *obyčejná diference* v bodě x_j.

Např.:

$$\begin{aligned}
\Delta^1 f_0 &= -f(x_0) + f(x_0+h) = -f_0 + f_1\\
\Delta^2 f_0 &= f(x_0) - 2f(x_0+h) + f(x_0+2h) = f_0 - 2f_1 + f_2\\
\Delta^3 f_0 &= -f_0 + 3f_1 - 3f_2 + f_3
\end{aligned}$$

atd.

Obdobným způsobem jako pro poměrné diference lze i pro obyčejné diference dokázat rekurentní vztah:

$$\Delta^{k+1} f_i = \Delta\left(\Delta^k f_i\right) = \Delta^k f_{i+1} - \Delta^k f_i. \tag{6.24}$$

Lemma. *Na ekvidistantní množině uzlů* $\{x_i\}_{i=0}^n$ *platí*

$$f[x_0, \ldots, x_n] = \frac{\Delta^n f_0}{n!\, h^n}. \tag{6.25}$$

Důkaz lze provést matematickou indukcí — viz cvičení.

Některé další vlastnosti poměrných a obyčejných diferencí lze nalézt např. v [23].

Věta 6.6. *Nechť uzly* x_i, $i = 0, 1, \ldots, n$ *jsou ekvidistantní,* $x_i = x_0 + ih$, $h > 0$. *Pak Newtonův interpolační polynom lze zapsat ve tvaru*

$$P_n(x_0 + th) = f_0 + \sum_{j=1}^{n} \frac{\Delta^j f_0}{j!} t(t-1)\ldots(t-j+1), \tag{6.26}$$

kde $x = x_0 + th$, t *je nová proměnná,* $t \in \mathbb{R}$.

Důkaz. Bod x, ve kterém počítáme hodnotu interpolačního polynomu, vyjádříme pomocí kroku h, $x = x_0 + th$, t je nová proměnná. Nyní

$$x - x_i = x_0 + th - (x_0 + ih) = (t-i)h.$$

Víme, že Newtonův polynom je tvaru

$$P_n(x) = f_0 + f[x_0, x_1](x - x_0) + \ldots + (x - x_0)\ldots(x - x_{n-1})f[x_0, \ldots, x_n].$$

Užitím vztahu (6.25) upravíme j-tý člen tohoto polynomu:

$$(x - x_0)\ldots(x - x_{j-1})f[x_0, \ldots, x_j] = (x - x_0)\ldots(x - x_{j-1})\frac{\Delta^j f_0}{j!\, h^j}$$

a dále, v důsledku toho, že $x - x_i = (t - i)h$, dostaneme

$$(x - x_0)\ldots(x - x_{j-1})f[x_0, \ldots, x_j] = \frac{t\ldots(t - j + 1)}{j!}\Delta^j f_0$$

a odtud plyne (6.26). □

Obdobným způsobem jako větu 6.6 lze dokázat následující větu 6.7.

Věta 6.7. *Nechť uzly x_i, $i = 0, 1, \ldots, n$, jsou ekvidistantní, $x_i = x_0 + ih$, $h > 0$. Pak Newtonův interpolační polynom lze zapsat ve tvaru*

$$P_n(x_n + sh) = f_n + \sum_{j=1}^{n} \frac{\Delta^j f_{n-j}}{j!} s(s + 1)\ldots(s + j - 1), \tag{6.27}$$

kde $x = x_n + sh$, $s \in \mathbb{R}$.

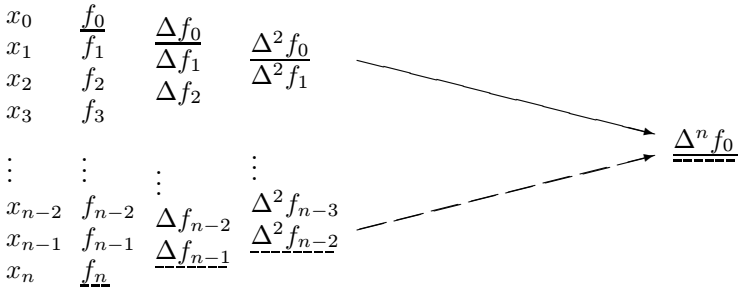
Důkaz. V tomto případě se bod x vyjádří pomocí bodu x_n: $x = x_n + sh$, $s \in R$ je nová proměnná a interpolační polynom sestrojíme v bodech $x_n, \ldots, x_0$, $x_{n-j} = x_n - jh$, tj.

$$P_n(x) = f_n + (x - x_n)f[x_n, x_{n-1}] + \ldots + (x - x_n)\ldots(x - x_1)f[x_n, \ldots, x_0].$$

□

Definice 6.5. Formule (6.26) se nazývá *Newtonův interpolační polynom pro interpolaci vpřed*. Formule (6.27) se nazývá *Newtonův interpolační polynom pro interpolaci vzad*.

Schematicky lze znázornit použití formulí vpřed a vzad takto:



Diference označené ⎯ se používají pro formuli vpřed, diference - - - se používají pro formuli vzad.

Poznámka 5. Z poněkud modifikované tabulky diferencí tzv. *Fraserova diagramu* lze odvodit celou řadu užitečných interpolačních formulí. Tento diagram lze najít např. v [19].

Zmíníme se nyní o minimalizaci chyby v případě, že uzly jsou ekvidistantní. Zřejmě má na velikost chyby rozhodující vliv chování funkce ω_{n+1}. Užitím substituce $x = x_0 + th$ lze funkci ω_{n+1} vyjádřit ve tvaru

$$\omega_{n+1}(x) = \omega_{n+1}(x_0 + th) = h^{n+1}t(t-1)\dots(t-n), \qquad t \in \mathbb{R}.$$

Nyní budeme vyšetřovat chování funkce

$$\varphi(t) = t(t-1)\dots(t-n). \tag{6.28}$$

pro $t \in [0, n]$. Nejdříve si všimněme, že funkce φ je lichá nebo sudá (v závislosti na n) vzhledem k bodu $(\frac{n}{2}, 0)$. Tento fakt plyne ze vztahu

$$\varphi(t) = (-1)^{n+1}\varphi(n-t),$$

tedy φ je sudá, je-li n liché, a lichá, je-li n sudé. Dále

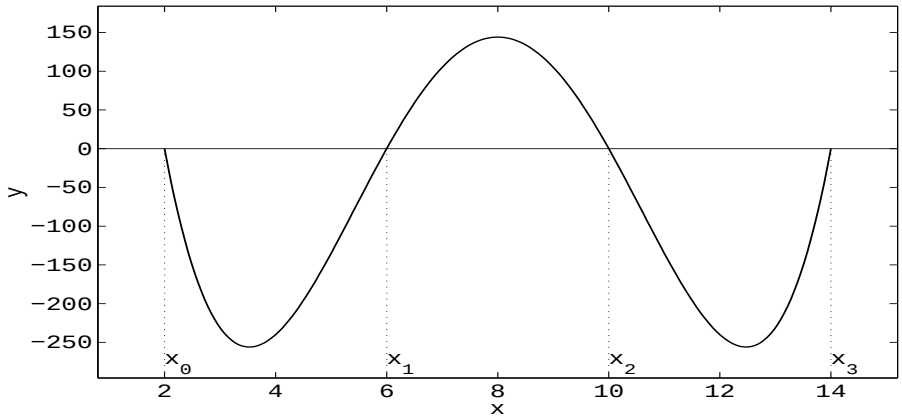
$$\varphi(t+1) = (t+1)t(t-1)\dots(t+1-n) = \frac{t+1}{t-n}\varphi(t).$$

Odtud plyne, že na intervalu $[i, i+1]$, $i = 0, \dots, n-1$, lze hodnoty funkce φ získat pomocí hodnot funkce na intervalu $[i-1, i]$ vynásobených faktorem $(t+1)/(t-n)$. Ovšem tento faktor je vždy záporný pro $t < n$. To znamená, že znaménka funkce φ se budou střídat při přechodu z jednoho intervalu na interval následující. Navíc

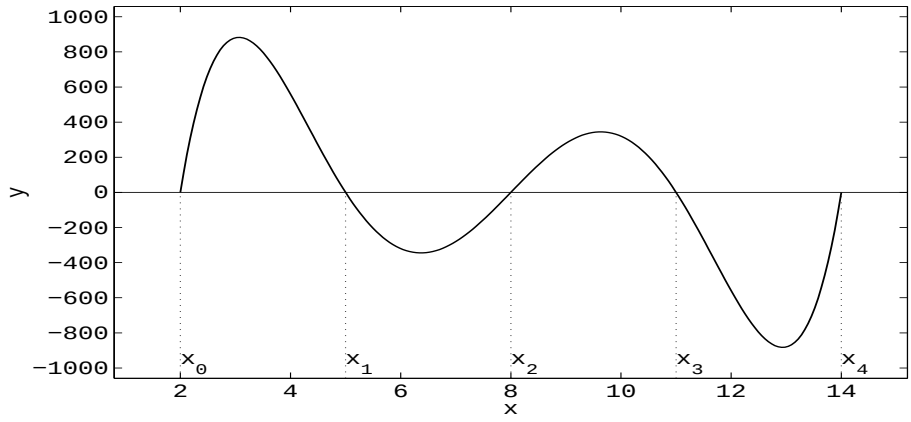
$$\left|\frac{t+1}{t-n}\right| < 1$$

pro $t \in [0, \frac{n-1}{2})$. Tedy, extremální hodnoty funkce φ na jednotlivých intervalech $[i, i+1]$ budou v absolutní hodnotě klesat až do středu intervalu $[0, n]$, a pak v důsledku symetrie opět porostou. Vně intervalu $[0, n]$ funkce φ v absolutní hodnotě velmi rychle roste. Tato skutečnost je ilustrována na následujících dvou případech — obr. 6.3 a 6.4 ilustrují průběh funkcí ω_4 a ω_5. Z uvedených poznatků plyne následující závěr: Veličina $|\omega_{n+1}(\bar{x})|$ bude minimální, a tedy i chyba interpolace bude minimální, jestliže pro interpolaci v bodě $\bar{x}$ vybereme $n+1$ uzlů nejbližších bodu $\bar{x}$. V případě, že bod $\bar{x}$ leží na začátku tabulky, je vhodné užít Newtonovu interpolační formuli vpřed, leží-li bod $\bar{x}$ blízko konce tabulky, je vhodná Newtonova interpolační formule vzad. Pro hodnoty $\bar{x}$ ležící vně intervalu $[x_0, x_n]$ lze očekávat značnou chybu interpolace. V tomto případě hovoříme o *extrapolaci*.

Příklad 6.4. V následující tabulce jsou dány hodnoty Besselovy funkce 1. druhu řádu nula. Aproximujte hodnotu této funkce v bodě $\bar{x} = 1{,}5$. Užijte Lagrangeových polynomů různých stupňů a porovnejte výsledky s přesnou hodnotou.



Obr. 6.3: Polynom $\omega_4(x) = (x-2)(x-6)(x-10)(x-14)$



Obr. 6.4: Polynom $\omega_5(x) = (x-2)(x-5)(x-8)(x-11)(x-14)$

x_i	1,0	1,3	1,6	1,9	2,2
$f(x_i)$	0,7651977	0,6200860	0,4554022	0,2818186	0,1103623

Řešení:

a) Lineární interpolace: bod $\bar{x} = 1{,}5$ leží mezi $x_1 = 1{,}3$ a $x_2 = 1{,}6$, zvolíme za uzly interpolace 1,3; 1,6.

$$P_1(1{,}5) = \frac{(1{,}5 - 1{,}6)}{(1{,}3 - 1{,}6)} \cdot 0{,}6200860 + \frac{1{,}5 - 1{,}6}{1{,}6 - 1{,}3} \cdot 0{,}4554022 = 0{,}5102968$$

b) Pro aproximaci polynomem druhého stupně zvolíme uzly 1,6; 1,3; 1,9. Dostaneme přibližnou hodnotu

$$P_2(1{,}5) = 0{,}5112857.$$

c) Pro polynom třetího stupně zvolíme uzly 1,6; 1,3; 1,9; 1,0 a dostaneme hodnotu

$$P_3(1{,}5) = 0{,}5118127.$$

d) Pro polynom čtvrtého stupně užijeme všech uzlů a výsledná hodnota je

$$P_4(1{,}5) = 0{,}5118200.$$

e) Pro polynom prvního stupně užijeme nyní uzlů $x_0 = 1$, $x_1 = 1{,}3$. Je

$$P_1(1{,}5) = 1{,}17752595.$$

f) Pro polynom druhého stupně užijeme nyní uzlů $x_0 = 1{,}6$, $x_1 = 1{,}9$, $x_2 = 2{,}2$. Je

$$P_2(1{,}5) = 1{,}73385945.$$

Porovnáme vypočtené hodnoty s přesnou hodnotou $f(1{,}5) = 0{,}5118277$:

a) $|P_1(1{,}5) - f(1{,}5)| \approx 1{,}53.10^{-3}$

b) $|P_2(1{,}5) - f(1{,}5)| \approx 5{,}42.10^{-4}$

c) $|P_3(1{,}5) - f(1{,}5)| \approx 1{,}5.10^{-5}$

d) $|P_4(1{,}5) - f(1{,}5)| \approx 7{,}7.10^{-6}$

e) $|P_1(1{,}5) - f(1{,}5)| \approx 6{,}6.10^{-1}$

f) $|P_2(1{,}5) - f(1{,}5)| \approx 1{,}22$

Je jasné, že v případech e), f) je chyba (vzhledem k extrapolaci) podstatně větší. Přenecháváme čtenáři, aby se pokusil navrhnout algoritmus pro konstrukci takové posloupnosti polynomů. Nelze však očekávat, že ve všech případech se bude s rostoucím počtem uzlů zvyšovat také přesnost aproximace (viz odstavec 6.4).

Na závěr tohoto odstavce ukážeme zajímavý příklad použití interpolačního polynomu.

Příklad 6.5. Dokažte, že platí

$$\sum_{k=0}^{n-1}(-1)^k\frac{n-k}{n}\binom{n+m}{k}=(-1)^{n-1}\frac{(n+m-2)!}{(m-1)!\,n!}\,.$$

Řešení: Nechť

$$f(x)=\frac{(n-x)(n-1-x)\dots(2-x)}{n!}$$

a sestrojme interpolační polynom P_{n-1} stupně $n-1$ pro ekvidistantní uzly dané hodnotami $x_0=0$, $h=1$:

$$\begin{aligned}P_{n-1}(x) &= \sum_{k=0}^{n-1}\frac{\Delta^k f_0}{h^k}\frac{(x-x_0)(x-x_0-h)\dots(x-x_0-(k-1)h)}{k!}=\\ &= \sum_{k=0}^{n-1}\Delta^k f_0\frac{x(x-1)\dots(x-(k-1))}{k!}\,.\end{aligned}$$

Platí

$$P_{n-1}(x_0)=P_{n-1}(0)=f(0)=1,\quad P_{n-1}(x_0+h)=P_{n-1}(1)=f(1)=\frac{1}{n}$$

a dále

$$P_{n-1}(x_0+kh)=P_{n-1}(k)=f(k)=0,\quad \text{pro } k=2,\dots,n-1\,.$$

Odtud

$$\Delta^k f_0=\sum_{r=0}^{k}(-1)^{k-r}\binom{k}{r}f(x_0+rh)=(-1)^k\frac{n-k}{n}\,,$$

a protože funkce f je polynomem stupně $n-1$, musí být totožná s polynomem P_{n-1}, tedy

$$\frac{(n-x)(n-1-x)\dots(2-x)}{n!}=\sum_{k=0}^{n-1}(-1)^k\frac{n-k}{n}\frac{x(x-1)\dots(x-(k-1))}{k!}\,.$$

Pro $x=n+m$ dostaneme požadovanou formuli

$$\sum_{k=0}^{n-1}(-1)^k\frac{n-k}{n}\binom{n+m}{k}=(-1)^{n-1}\frac{(n+m-2)!}{(m-1)!\,n!}\,,$$

nebo další transformací (výměnou role n a m a položíme-li $i=m-k$)

$$\sum_{i=1}^{m}(-1)^{m-1}\frac{i}{m}\binom{n+m}{n+i}=(-1)^{m-1}\frac{(n+m-2)!}{(n-1)!\,m!}\,.$$

§ 6.4. Obecný interpolační proces

Uvažujme nyní následující problém: V intervalu $[a, b]$ vybereme uzly tvořící nekonečnou trojúhelníkovou matici:

$$\begin{array}{ccccccc} x_0^{(0)} & & & & & & \\ x_0^{(1)} & x_1^{(1)} & & & & & \\ x_0^{(2)} & x_1^{(2)} & x_2^{(2)} & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ x_0^{(n)} & x_1^{(n)} & x_2^{(n)} & \cdots & x_n^{(n)} & & \\ \vdots & \vdots & \vdots & & & \ddots & \end{array} \tag{6.29}$$

Pro danou funkci sestrojíme posloupnost Lagrangeových interpolačních polynomů P_n tak, že k sestrojení P_n užijeme $(n+1)$-ho řádku matice (6.29), tj.

$$P_n(x_i^{(n)}) = f(x_i^{(n)}), \qquad i = 0, 1, \ldots, n.$$

Ptáme se: *Bude posloupnost* $\{P_n\}$ *konvergovat stejnoměrně k funkci* f *na intervalu* $[a, b]$*?* Dá se ukázat, že pro každou matici (6.29) existuje třída funkcí, pro niž platí stejnoměrná konvergence, ale tato třída je podstatně užší než $C[a, b]$.

Věta 6.8. (G. Fáber). *Pro každou matici* (6.29) *existuje spojitá funkce, pro kterou příslušná posloupnost interpolačních polynomů nekonverguje stejnoměrně k* f *na intervalu* $[a, b]$.

Důkaz je uveden v [16].

Ve Fáberově větě se mluví o neexistenci stejnoměrné konvergence posloupností P_n k f. Není ale vyloučeno, že v některých bodech konvergují polynomy P_n k funkci f. Následující příklad ilustruje možnost divergence interpolačního procesu v jednotlivých bodech.

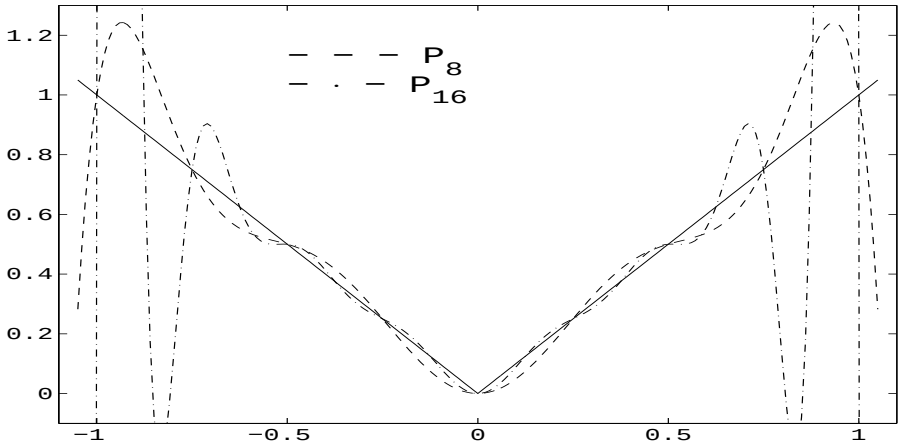
Věta 6.9. (S. N. Bernštejn). *Interpolační polynom* P_n *sestrojený pro funkci* $|x|$ *na ekvidistantní množině uzlů intervalu* $[-1, 1]$ *(tak, že* $x_0 = -1$*,* $x_n = 1$*) nekonverguje s rostoucím* n *k* $|x|$ *ani v jednom bodě intervalu* $[-1, 1]$ *různém od bodů* $-1, 0, 1$.

Důkaz je opět uveden v [16], příklad je ilustrován na obr. 6.5.

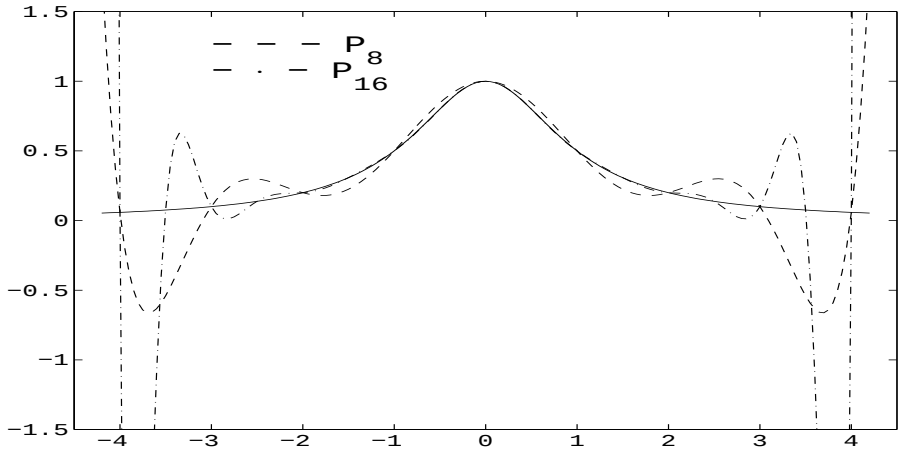
Zajímavý příklad — příklad Rungeho — je uveden v [8]. Tento příklad se týká interpolace funkce $f(x) = 1/(1+x^2)$, $x \in [-4, 4]$, a interpolační proces je ilustrován na obr. 6.6.

Na druhé straně platí:

Věta 6.10. (I. Marcinkiewicz). *Pro každou spojitou funkci* f *existuje taková matice* (6.29)*, že odpovídající posloupnost interpolačních polynomů konverguje stejnoměrně k funkci* f *na intervalu* $[a, b]$.



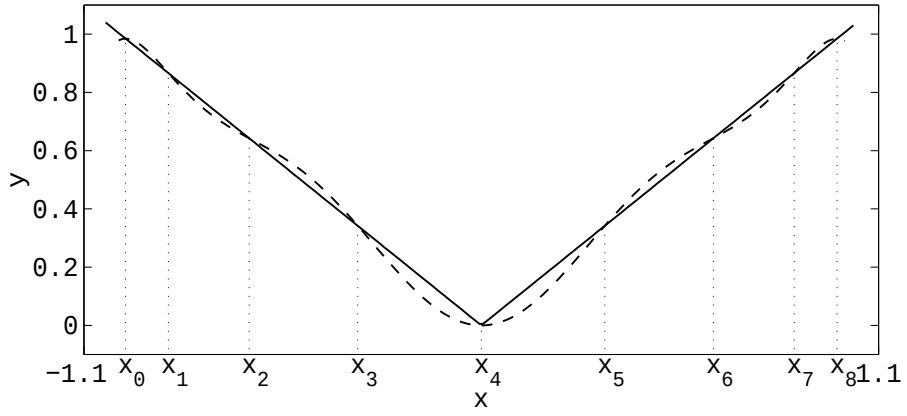
Obr. 6.5: Interpolační polynomy pro $f(x) = |x|$ na ekvidistantních uzlech



Obr. 6.6: Interpolační polynomy pro $f(x) = 1/(1 + x^2)$ na ekvidistantních uzlech

Důkaz lze opět nalézt v [16], kde je obecnému interpolačnímu procesu věnována značná pozornost.

Poznámka 6. Zvolíme-li v předchozích příkladech za uzly kořeny Čebyševových polynomů definovaných vztahem (6.18), dostaneme konvergentní proces. Chování posloupností odpovídajících interpolačních polynomů je zřejmé z obr. 6.7, 6.8.



Obr. 6.7: Interpolační polynom $P_8 \in \Pi_8$ pro $f(x) = |x|$ na Čebyševových uzlech

§ 6.5. Iterovaná interpolace

Interpolační problém nemusíme řešit ihned jako celek pro všechny dané uzly, ale můžeme začít s menším počtem uzlů a postupně zkonstruovat celý interpolační polynom. Tímto problémem se budeme nyní zabývat.

Pro danou množinu bodů $\{(x_i, f_i);\ i = 0, 1, \ldots, n\}$, označme

$$P_{i_0 i_1 \ldots i_k} \in \Pi_k, \quad k \leq n$$

takový polynom, pro který

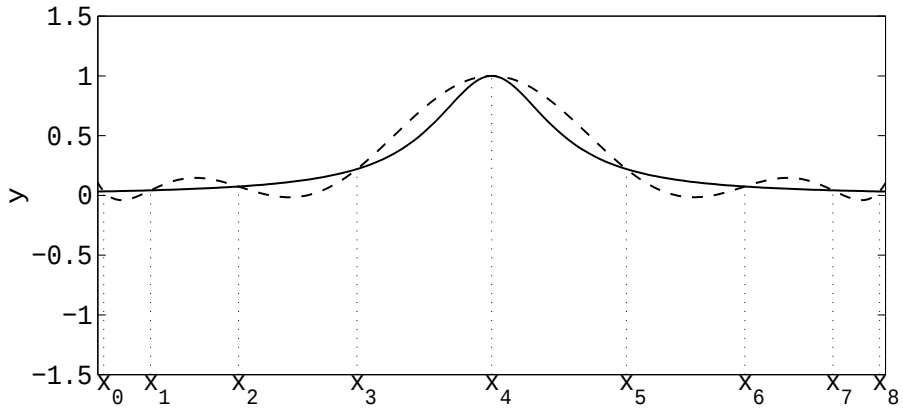
$$P_{i_0 i_1 \ldots i_k}(x_{i_j}) = f_{i_j}, \qquad j = 0, 1, \ldots, k.$$

Lemma. *Platí*

$$P_{i_j}(x) = f_{i_j}, \qquad j = 0, 1, \ldots, k, \tag{6.30}$$

$$P_{i_0 \ldots i_k}(x) = \frac{1}{x_{i_k} - x_{i_0}} \begin{vmatrix} P_{i_1 \ldots i_k}(x) & x - x_{i_k} \\ P_{i_0 \ldots i_{k-1}}(x) & x - x_{i_0} \end{vmatrix}. \tag{6.31}$$

Tento postup nazýváme *iterovanou interpolací.*



Obr. 6.8: Interpolační polynom $P_8 \in \Pi_8$ pro $f(x) = 1/(1+x^2)$ na Čebyševových uzlech

Důkaz. Vztah (6.30) je zřejmý. Abychom dokázali (6.31), označíme pravou stranu $R(x)$ a ukážeme, že má charakteristické vlastnosti interpolačního polynomu. Zřejmě R je polynom stupně nejvýše k. Podle definice $P_{i_0\dots i_{k-1}}$, $P_{i_1\dots i_k}$ je

$$\begin{aligned} R(x_{i_0}) &= P_{i_0\dots i_{k-1}}(x_{i_0}) = f_{i_0}, \\ R(x_{i_k}) &= P_{i_1\dots i_k}(x_{i_k}) = f_{i_k}, \\ R(x_{i_j}) &= \frac{(x_{i_j}-x_{i_0})f_{i_j} - (x_{i_j}-x_{i_k})f_{i_j}}{x_{i_k}-x_{i_0}} = f_{i_j} \end{aligned}$$

pro $j = 1, \dots, k-1$. Tedy $R \equiv P_{i_0\dots i_k}$, což plyne z jednoznačnosti interpolačního polynomu v daném bodě x. □

Podle vzorce (6.31) lze výpočet uspořádat například takto:

$$\begin{array}{cccccc} & k=0 & 1 & 2 & 3 \\ \hline x_0 & f_0 = P_0(x) & & & \\ x_1 & f_1 = P_1(x) & P_{01}(x) & & \\ x_2 & f_2 = P_2(x) & P_{12}(x) & P_{012}(x) & \\ x_3 & f_3 = P_3(x) & P_{23}(x) & P_{123}(x) & P_{0123}(x) \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \tag{6.32}$$

První sloupec tabulky obsahuje předepsané hodnoty f_i. Hodnoty v dalších sloupcích počítáme podle vztahu (6.31). Je například

$$P_{012}(x) = \frac{(x-x_0)P_{12} - (x-x_2)P_{01}}{x_2 - x_0}.$$

Uvedený algoritmus (6.32) se nazývá *Nevillův algoritmus*.

Příklad 6.6. Je dána tabulka hodnot funkce f. Užijte Nevillova schematu pro výpočet $f(1)$.

x_i	f_i	$P_{i,i+1}$	$P_{i,i+1,i+2}$	$P_{i,i+1,i+2,i+3}$
0	1			
2	3	2		
3	2	4	8/3	
5	5	−1	17/3	49/15

To znamená, že $f(1) \approx 49/15$.

Pro snazší použití na počítači zaveďme následující označení. Položme

$$T_{i,j} = P_{i-j,i-j+1,\ldots,i-1,i}.$$

Rekurentní vztahy (6.31) můžeme nyní zapsat takto

$$T_{i,j}(x) = \frac{(x - x_{i-j})T_{i,j-1}(x) - (x - x_i)T_{i-1,j-1}(x)}{x_i - x_{i-j}} \qquad \begin{array}{l} j = 1, 2, 3, \ldots \\ i = j, j+1, \ldots \end{array}$$

$$T_{i0} = f_i, \qquad i = 0, 1, \ldots, n$$

a pak

$$T_{nn} = P_{01\ldots n}.$$

Příslušná tabulka je tvaru

	$k = 0$	1	2	3
x_0	$f_0 = T_{00}$			
x_1	$f_1 = T_{10}$	T_{11}		
x_2	$f_2 = T_{20}$	T_{21}	T_{22}	
x_3	$f_3 = T_{30}$	T_{31}	T_{32}	T_{33}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

(6.33)

Jako kriteria pro zastavení výpočtu lze užít nerovnosti

$$|T_{i,i} - T_{i-1,i-1}| < \varepsilon,$$

kde ε je předepsaná přesnost. Jestliže tato nerovnost není splněna, přidá se další uzel x_{i+1}.

Závěrem tohoto odstavce připomeňme *Aitkenův algoritmus* ([11]). Ten je rovněž založen na vztazích (6.30), (6.31), ale používá jiných polynomů během výpočtu. Tabulka je následující:

	$k=0$	1	2	3
x_0	$f_0 = P_0$			
x_1	$f_1 = P_1$	P_{01}		
x_2	$f_2 = P_2$	P_{02}	P_{012}	
x_3	$f_3 = P_3$	P_{03}	P_{013}	P_{0123}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Příklad 6.7. Pro tabulku hodnot z příkladu 6.6 vypočtěte $f(1)$ pomocí Aitkenova schematu.

x_i	f_i	P_{0i}	P_{01i}	P_{012i}
0	1			
2	3	2		
3	2	4/3	8/3	
5	5	9/5	31/15	49/15

Tedy $f(1) \approx 49/15$.

Obě schemata dávají tentýž výsledek, neboť se jedná o užití interpolačního polynomu v uzlech $x_0 = 0$, $x_1 = 2$, $x_2 = 3$, $x_3 = 5$.

§ 6.6. Inverzní interpolace

Nyní ukážeme, jak lze užít interpolace k nalezení přibližného kořene funkce. Tento proces budeme nazývat *inverzní interpolací*. Předpokládejme, že funkce f, f' jsou spojité v $[a,b]$, $f'(x) \neq 0$ v $[a,b]$ a nechť ξ je kořen funkce f ležící v intervalu $[a,b]$. Nechť $x_i \in [a,b]$, $i = 0,1,\ldots,n$, $x_i \neq x_k$ pro $i \neq k$ a nechť $f(x_i) = y_i$, $i = 0,1,\ldots,n$. Naším úkolem je najít přibližnou hodnotu kořene ξ. Budeme postupovat takto: Sestrojíme interpolační polynom $P_n(y)$ pro funkci $f^{-1}(y)$ v bodech

$$(y_i, x_i), \qquad x_i = f^{-1}(y_i), \qquad i = 0,1,\ldots,n,$$

tj.

$$P_n(y_i) = f^{-1}(y_i), \qquad i = 0,1,\ldots,n.$$

Protože $0 = f(\xi)$, je $\xi = f^{-1}(0)$ a za přibližnou hodnotu kořene ξ lze vzít číslo $P_n(0)$. Pro výpočet lze s výhodou užít iterované interpolace.

§ 6.7. Sestavování tabulek

Nyní se zabývejme následujícím úkolem: Je třeba sestavit tabulku hodnot nějaké funkce tak, aby chyba při interpolaci hodnot funkce polynomem daného stupně m

nepřevýšila ε. V takovém případě říkáme, že tabulka *připouští interpolaci stupně* m. Tabulky, se kterými se většinou setkáváme, připouštějí lineární interpolaci. Budeme se zabývat tabulkami s ekvidistantními uzly. Otázka tedy je: Jak zvolit krok h, aby při lineární interpolaci byla chyba menší než ε? Nechť x je bod, ve kterém máme spočítat přibližnou hodnotu. Položme $x_0 < x < x_1$ a f aproximujeme interpolačním polynomem

$$f(x) = P_1(x) + E(x),$$

substituce $x = x_0 + th$, $t \in (0,1)$ je nová proměnná, vede ke vztahu

$$f(x_0 + th) = P_1(x_0 + th) + \frac{h^2 t(t-1)}{2} f''(\xi).$$

Protože $|t(t-1)| \leq \frac{1}{4}$, musí pro krok h platit

$$h^2 \max |f''(x)| \leq 8\varepsilon.$$

Příklad 6.8. Nechť je třeba sestavit tabulku funkce $\sin x$ na intervalu $[0, \frac{\pi}{2}]$ tak, aby chyba lineární interpolace byla menší než $0{,}5.10^{-6}$.
Řešení. Je

$$f(x) = \sin x \quad \Rightarrow \quad |f''(x)| \leq 1.$$

Z předchozího plyne

$$\begin{aligned} h^2 &\leq 4.10^{-6}, \\ h &\leq 2.10^{-3}. \end{aligned}$$

Přenecháváme čtenáři posoudit otázku přípustnosti *kvadratické interpolace*.

§ 6.8. Hermitova interpolace

Obecně lze předepsat hledanému polynomu nejen funkční hodnoty, ale také hodnoty derivací. Přesněji: Jsou dána reálná čísla x_i, $i = 0, 1, \ldots, m$, $x_i \neq x_k$ pro $i \neq k$, $f_i^{(k)}$, $k = 0, 1, \ldots, n_i - 1$, přičemž

$$\sum_{i=0}^{m} n_i = n + 1. \tag{6.34}$$

Hermitův interpolační problém spočívá v tom, že je třeba najít polynom $P_n \in \Pi_n$ takový, že

$$P_n^{(k)}(x_i) = f_i^{(k)}, \qquad i = 0, 1, \ldots, m; \qquad k = 0, 1, \ldots, n_i - 1. \tag{6.35}$$

V každém uzlu x_i je tedy předepsána nejen funkční hodnota, ale také hodnoty prvních $(n_i - 1)$ derivací.

Věta 6.11. *Pro daná reálná čísla x_i, $i = 0, 1, \ldots, m$, $x_i \neq x_k$ pro $i \neq k$, a hodnoty $f_i^{(k)}$, $k = 0, 1, \ldots, n_i - 1$, existuje právě jeden polynom $P_n \in \Pi_n$,*

$$\sum_{i=0}^{m} n_i = n + 1,$$

takový, že jsou splněny podmínky (6.35).

Obecný důkaz nebudeme provádět, ale uvedeme nyní konstrukci Hermitova interpolačního polynomu v jednodušším případě, kdy v každém uzlu x_i, $i = 0, 1, \ldots, m$, předepíšeme funkční hodnotu f_i a hodnotu první derivace f_i'. Je tedy třeba najít polynom $P_n \in \Pi_n$, $2m + 1 = n$, takový, že

$$\begin{aligned} P_{2m+1}(x_i) &= f_i, \qquad i \ = 0, 1, \ldots, m \\ P'_{2m+1}(x_i) &= f'_i, \qquad i \ = 0, 1, \ldots, m \end{aligned} \tag{6.36}$$

Polynom P_{2m+1} budeme hledat ve tvaru

$$P_{2m+1}(x) = \sum_{i=0}^{m} h_i(x) f_i + \sum_{i=0}^{m} \bar{h}_i(x) f'_i,$$

kde h_i, $\bar{h}_i$ jsou polynomy stupně $2m + 1$ pro $m > 0$. Polynomy h_i, $\bar{h}_i$ je třeba vybrat tak, aby byly splněny podmínky (6.36), tj.

$$\begin{aligned} h_i(x_j) &= \delta_{ij}, \qquad i, j = 0, 1, \ldots, m \\ h'_i(x_j) &= 0, \qquad i, j = 0, 1, \ldots, m \end{aligned} \tag{6.37}$$

$$\begin{aligned} \bar{h}_i(x_j) &= 0, \qquad i, j = 0, 1, \ldots, m \\ \bar{h}'_i(x_j) &= \delta_{ij}, \qquad i, j = 0, 1, \ldots, m\,, \end{aligned} \tag{6.38}$$

kde $\delta_{ij} = 0$ pro $i \neq j$, $\delta_{ij} = 1$ pro $i = j$ (Kroneckerův symbol). Sestrojme nejdříve polynomy h_i: Polynom h_i je polynom stupně $2m+1$, který má podle (6.37) kořeny $x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m$, přičemž jsou všechny tyto kořeny dvojnásobné, tj. $2m$ kořenů. Protože h_i je polynom stupně $2m + 1$, plyne odtud, že je tvaru

$$h_i(x) = t_i(x)(x - x_0)^2 \ldots (x - x_{i-1})^2 (x - x_{i+1})^2 \ldots (x - x_m)^2,$$

kde t_i je polynom stupně prvního. Bez újmy na obecnosti lze psát

$$\begin{aligned} h_i(x) &= u_i(x) \frac{(x - x_0)^2 \ldots (x - x_{i-1})^2 (x - x_{i+1})^2 \ldots (x - x_m)^2}{(x_i - x_0)^2 \ldots (x_i - x_{i-1})^2 (x_i - x_{i+1})^2 \ldots (x_i - x_m)^2} = \\ &= u_i(x) l_i^2(x), \end{aligned}$$

kde l_i je fundamentální polynom odvozený při konstrukci Lagrangeova interpolačního polynomu a u_i je lineární polynom: $u_i(x) = a_i x + b_i$, $i = 0, 1, \ldots, m$.

Koeficienty a_i, b_i určíme z podmínek

$$\begin{aligned} & & h_i(x_i) = 1, & & h_i'(x_i) = 0, \\ \text{tj.} & & u_i(x_i)l_i^2(x_i) = 1 & \Rightarrow & u_i(x_i) = 1, \\ & u_i'(x)l_i^2(x) + 2u_i(x)l_i(x)l_i'(x) = 0 & \Rightarrow & & u_i'(x_i) + 2u_i(x_i)l_i'(x_i) = 0. \end{aligned}$$

Odtud

$$a_i = -2l_i'(x_i), \qquad b_i = 1 + 2x_i l_i'(x_i),$$

polynom u_i je tvaru

$$u_i(x) = 1 - 2(x - x_i)l_i'(x_i), \qquad i = 0, 1, \ldots, m$$

a hledaný polynom h_i:

$$h_i(x) = (1 - 2(x - x_i)l_i'(x_i))\, l_i^2(x), \qquad i = 0, 1, \ldots, m\,. \tag{6.39}$$

Obdobně sestrojíme polynomy $\bar{h}_i$: $\bar{h}_i$ je polynom stupně $2m+1$ a má podle (6.38) kořeny $x_0, x_1, \ldots, x_m$, přičemž kořeny $x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m$ jsou dvojnásobné, tj. celkem $2m+1$ kořenů. Předpokládaný tvar polynomu $\bar{h}_i$ je

$$\bar{h}_i(x) = A_i(x - x_i)(x - x_0)^2 \ldots (x - x_{i-1})^2(x - x_{i+1})^2 \ldots (x - x_m)^2,$$

což lze opět bez újmy na obecnosti zapsat ve tvaru

$$\bar{h}_i(x) = B_i(x - x_i)l_i^2(x)$$

a hodnotu B_i určíme tak, aby $\bar{h}_i'(x_i) = 1$, tj.

$$B_i = 1.$$

Polynomy $\bar{h}_i$ jsou tvaru

$$\bar{h}_i(x) = (x - x_i)l_i^2(x), \qquad i = 0, 1, \ldots, m. \tag{6.40}$$

Hermitův polynom pro dané hodnoty (6.36) má tvar

$$P_{2m+1}(x) = \sum_{i=0}^{m} h_i(x)f_i + \sum_{i=0}^{m} \bar{h}_i(x)f_i'$$

s polynomy h_i, $\bar{h}_i$ danými vztahy (6.39), (6.40).

Poznámka 7. Pro $m = 0$ dostáváme nejjednodušší Hermitův polynom, který je prvního stupně a je daný vztahem

$$P_1(x) = f(a) + (x - a)f'(a)\,.$$

Pro určení tohoto polynomu potřebujeme dvě podmínky – hodnotu funkce a derivace v daném bodě.

Příklad 6.9. Najděte Hermitův interpolační polynom, je-li dáno

x_i	0	1	4
f_i	2	5	1
f_i'	1	-1	2

Polynomy l_i jsou tvaru

$$l_0(x) = \tfrac{1}{4}(x-1)(x-4), \qquad l_1(x) = -\tfrac{1}{3}x(x-4), \qquad l_2(x) = \tfrac{1}{12}x(x-1)$$

Podle (6.39), (6.40) dostaneme pro h_i a $\bar{h}_i$

$$\begin{aligned} h_0(x) &= \tfrac{1}{32}(x-1)^2(x-4)^2(2+5x), & \bar{h}_0(x) &= \tfrac{1}{16}x(x-1)^2(x-4)^2 \\ h_1(x) &= \tfrac{1}{27}x^2(x-4)^2(7-4x), & \bar{h}_1(x) &= \tfrac{1}{9}x^2(x-1)(x-4)^2 \\ h_2(x) &= \tfrac{1}{864}x^2(x-1)^2(34-7x), & \bar{h}_2(x) &= \tfrac{1}{144}(x-4)x^2(x-1)^2 \end{aligned}$$

Hledaný polynom $P_5 \in \Pi_5$ je tvaru

$$\begin{aligned} P_5(x) = {} & \tfrac{1}{16}(x-1)^2(x-4)^2(2+5x) + \tfrac{5}{27}x^2(x-4)^2(7-4x) + \\ & + \tfrac{1}{864}x^2(x-1)^2(34-7x) + \tfrac{1}{16}x(x-1)^2(x-4)^2 - \\ & - \tfrac{1}{9}(x-1)x^2(x-4)^2 + \tfrac{1}{72}(x-4)x^2(x-1)^2. \end{aligned}$$

Příklad 6.10. Pro funkci $f(x) = x + \sin x$ sestrojte Hermitův interpolační polynom splňující v uzlech $x_0 = 5{,}5$, $x_1 = 12{,}5$, $x_2 = 3$ podmínky (6.36). Je zřejmé, že $P_5 \in \Pi_5$. Na obr. 6.9 jsou grafy polynomů h_i, $\bar{h}_i$, $i = 0, 1, 2$, obr. 6.10 znázorňuje graf funkce f (plná čára) a graf příslušného polynomu P_5 (čárkovaně).

Chyba interpolace pro Hermitův interpolační polynom může být odvozena stejným způsobem jako při Lagrangeově interpolaci. Odpovídající větu lze formulovat takto:

Věta 6.12. *Nechť funkce f je $(n+1)$-krát diferencovatelná v intervalu $[a, b]$ a nechť jsou dány body $x_i \in [a, b]$, $i = 0, 1, \ldots, m$, $x_i \neq x_k$ pro $i \neq k$. Jestliže polynom $P_n \in \Pi_n$,*

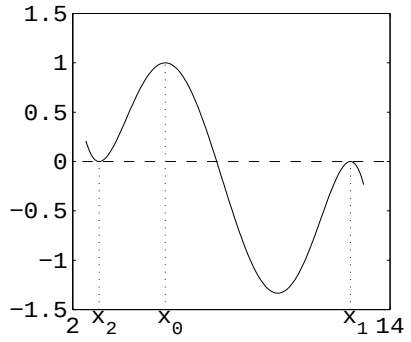
$$\sum_{i=0}^{m} n_i = n + 1,$$

splňuje interpolační podmínky

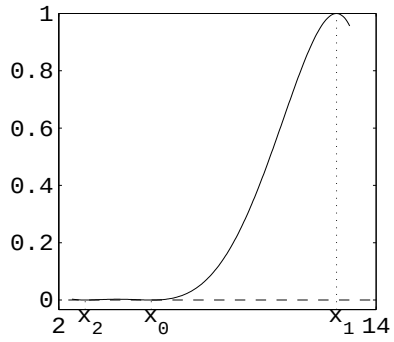
$$P_n^{(k)}(x_i) = f^{(k)}(x_i), \qquad k = 0, 1, \ldots, n_i - 1; \qquad i = 0, 1, \ldots, m,$$

pak ke každému $\bar{x} \in [a, b]$ existuje $\bar{\xi} \in (a, b)$ tak, že

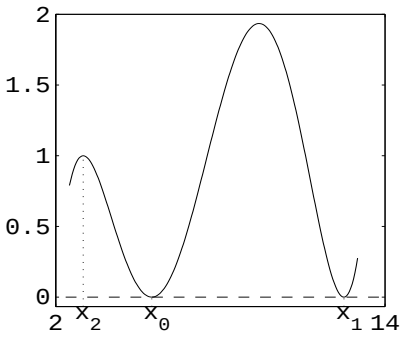
$$f(\bar{x}) - P_n(\bar{x}) = \frac{\omega_{n+1}(\bar{x})}{(n+1)!} f^{(n+1)}(\bar{\xi}), \qquad \bar{\xi} = \bar{\xi}(\bar{x}), \tag{6.41}$$



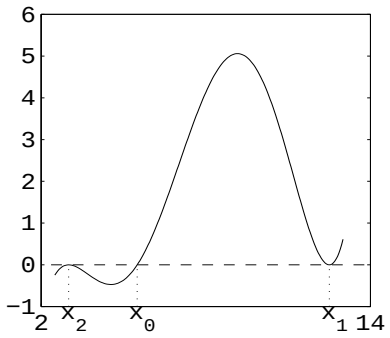
Fundamentální polynom $h_0(x)$



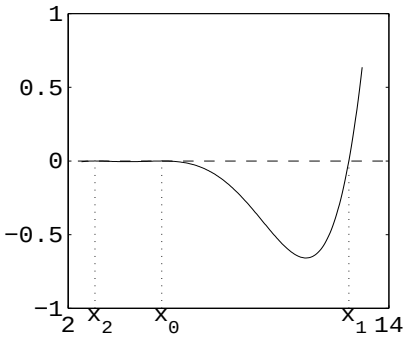
Fundamentální polynom $h_1(x)$



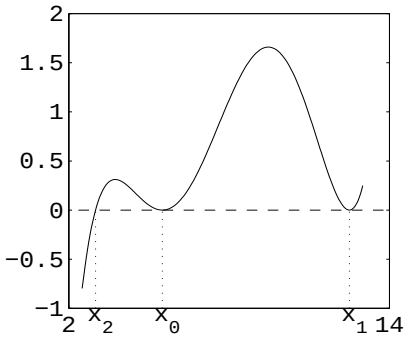
Fundamentální polynom $h_2(x)$



Fundamentální polynom $\bar{h}_0(x)$

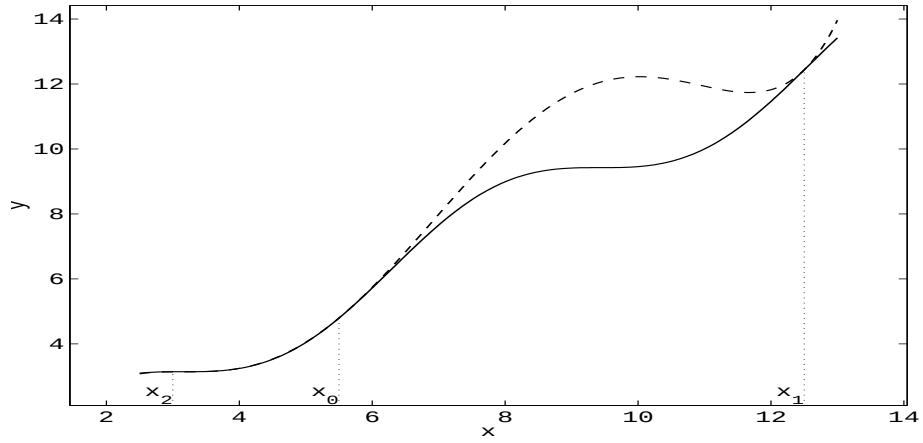


Fundamentální polynom $\bar{h}_1(x)$



Fundamentální polynom $\bar{h}_2(x)$

Obr. 6.9: Fundamentální polynomy h_i, $\bar{h}_i$



Obr. 6.10: Hermitův interpolační polynom $P_5 \in \Pi_5$ pro funkci $f(x) = x + \sin x$

kde

$$\omega_{n+1}(x) = (x - x_0)^{n_0} \dots (x - x_m)^{n_m}.$$

Poznámka 8. Odvození tvaru Hermitova interpolačního polynomu v obecném případě lze najít např. v [2], [19]. Pro konstrukci Hermitova interpolačního polynomu lze také s výhodou užít Lagrangeova polynomu. Postup je následující: Zapišme Hermitův polynom $P_n \in \Pi_n$ pro hodnoty (6.35) ve tvaru

$$P_n(x) = P_m(x) + \omega_{m+1}(x) H_{n-m-1}(x),$$

kde $P_m \in \Pi_m$ je Lagrangeův polynom pro hodnoty $(x_i, f_i^{(0)})$, $i = 0, 1, \dots, m$, $\omega_{m+1} = (x - x_0) \dots (x - x_m)$, $H_{n-m-1} \in \Pi_{n-m-1}$ je polynom stupně nejvýše $n - m - 1$ a určíme jej ze zbývajících $(n - m)$ podmínek, tj. z podmínek

$$P_n^{(k)}(x_i) = f_i^{(k)}, \qquad k = 1, \dots, n_i - 1, \qquad i = 0, 1, \dots, m. \tag{6.42}$$

Podmínkami (6.42) je polynom H_{n-m-1} určen jednoznačně.

Příklad 6.11. Najděte Hermitův interpolační polynom, je-li dáno:

x_i	0	1	2
f_i	1	-1	0
f_i'	0	0	0
f_i''	0		

Řešení: Budeme hledat polynom $P_6 \in \Pi_6$ ve tvaru

$$P_6(x) = P_2(x) + \omega_3(x) H_3(x)\,,$$

kde $\omega_3(x) = x(x-1)(x-2)$ a polynom $H_3 \in \Pi_3$, $H_3(x) = ax^3 + bx^2 + cx + d$, sestrojíme tak, aby platilo $P_6'(0) = 0$, $P_6'(1) = 0$, $P_6'(2) = 0$, $P_6''(0) = 0$. P_2 je Lagrangeův interpolační polynom pro body $(0,1)$, $(1,-1)$, $(2,0)$ a je zřejmě tvaru

$$P_2(x) = \frac{(3x-1)(x-2)}{2}\,.$$

Pak

$$P_6(x) = \frac{(3x-1)(x-2)}{2} + x(x-1)(x-2)H_3(x)\,.$$

Počítejme první a druhou derivaci polynomu P_6:

$$\begin{aligned} P_6'(x) &= \frac{6x-7}{2} + (3x^2-6x+2)H_3(x) + x(x-1)(x-2)H_3'(x) \\ P_6''(x) &= 3 + (6x-6)H_3(x) + 2(3x^2-6x+2)H_3'(x) + x(x-1)(x-2)H_3''(x) \end{aligned}$$

Z interpolačních podmínek pro derivace nyní plyne:

$$\begin{aligned} P_6'(0) = 0 &\;\Rightarrow\; 0 = -\frac{7}{2} + 2H_3(0) \\ P_6'(1) = 0 &\;\Rightarrow\; 0 = -\frac{1}{2} - H_3(1) \\ P_6'(2) = 0 &\;\Rightarrow\; 0 = \frac{5}{2} + 2H_3(2) \\ P_6''(0) = 0 &\;\Rightarrow\; 0 = 3 + 4H'3(0) - 6H_3(0) \end{aligned}$$

Odtud

$$\begin{aligned} H_3(0) = \frac{7}{4} &\;\Rightarrow\; d = \frac{7}{4} \\ H_3(1) = -\frac{1}{2} &\;\Rightarrow\; a+b+c+d = -\frac{1}{2} \\ H_3(2) = -\frac{5}{4} &\;\Rightarrow\; 8a+4b+2c+d = -\frac{5}{4} \\ -4H_3'(0) + 6H_3(0) = 3 &\;\Rightarrow\; -4c+6d = 3 \end{aligned}$$

Vypočteme

$$d = \frac{7}{4} \quad c = \frac{15}{8} \quad b = -\frac{105}{16} \quad a = \frac{39}{16}\,,$$

tedy

$$H_3(x) = \frac{1}{16}(39x^3 - 105x^2 + 30x + 28)\,.$$

Výsledný Hermitův polynom je

$$P_6(x) = \frac{(2x-1)(x-2)}{2} + \frac{1}{16}x(x-1)(x-2)(39x^3 - 105x^2 + 30x + 28)\,.$$

S Hermitovým interpolačním polynomem jsme se už vlastně setkali v matematické analýze. Připomeňme si Taylorův vzorec tvaru

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2}f''(a) + \dots$$
$$\dots + \frac{(x-a)^n}{n!}f^{(n)}(a) + \frac{(x-a)^{n+1}}{(n+1)!}f^{(n+1)}(\xi),$$

kde ξ leží v intervalu určeném body a, x, tj.

$$f(x) = P_n(x) + \frac{(x-a)^{n+1}}{(n+1)!}f^{(n+1)}(\xi)$$

a $P_n \in \Pi_n$ je v podstatě Hermitův interpolační polynom v případě, že je dán pouze jeden uzel $x_0 = a$ a požadujeme v tomto uzlu rovnost derivací až do řádu n včetně.

Z předchozích úvah plyne, že výpočet tvaru Hermitova interpolačního polynomu může být časově dosti náročný. Popíšeme alternativní postup založený na Newtonově tvaru interpolačního polynomu (viz např. [2])

Předpokládejme, že jsou dány hodnoty $(x_i, f^{(k)}(x_i))$, $i = 0, \dots, m$, $k = 0, \dots, n_i - 1$ a nechť $\sum_{i=0}^{m} n_i = n+1$. Předpokládejme dále, že $f \in C^{n+1}[a,b]$, $x_i \in [a,b]$, $i = 0, \dots, m$.

Definujme poměrné diference s násobnými uzly vztahem

$$f[\underbrace{x_j, \dots, x_j}_{n_j}] = \frac{f^{(n_j-1)}(x_j)}{(n_j-1)!}, \quad j = 0, \dots, m$$

a obecně

$$f[\underbrace{x_0, \dots, x_0}_{n_0}, \underbrace{x_1, \dots, x_1}_{n_1}, \dots, \underbrace{x_m, \dots, x_m}_{n_m}] =$$
$$= \frac{1}{x_m - x_0}\Big(f[\underbrace{x_0, \dots, x_0}_{n_0-1}, \underbrace{x_1, \dots, x_1}_{n_1}, \dots, \underbrace{x_m, \dots, x_m}_{n_m}] -$$
$$- f[\underbrace{x_0, \dots, x_0}_{n_0}, \underbrace{x_1, \dots, x_1}_{n_1}, \dots, \underbrace{x_m, \dots, x_m}_{n_m-1}]\Big).$$

Sestavení tabulky diferencí.

Sloupec 1: uzly interpolace, přičemž každý uzel opakujeme tolikrát, jaká je jeho násobnost, t.j. uzel x_i n_i-krát, $i = 0, \dots, m$.

Sloupec 2: odpovídající funkční hodnoty $f^{(0)(x_i)}$, $i = 0, \dots, m$.

Sloupec 3: poměrné diference, jsou-li sousední uzly různé, jinak první derivace

Sloupec 4: opět poměrné diference, jsou-li sousední uzly různé, jinak druhé derivace dělené 2!

$\vdots$

Sloupec $(n+2)$ **(poslední):** poměrná diference $f[\underbrace{x_0,\dots,x_0}_{n_0},\dots,\underbrace{x_m,\dots,x_m}_{n_m}]$

Tabulka poměrných diferencí.

$$
\begin{array}{cccccccc}
n_0 \left\{\begin{array}{cc} x_0 & f_0 \\ x_0 & f_0 \\ x_0 & f_0 \\ x_0 & f_0 \\ \vdots & \vdots \end{array}\right. & & & & & & & \\
 & & > & f'(x_0) & > & \frac{f''(x_0)}{2!} & > & \frac{f'''(x_0)}{3!} \quad \dots \\
 & & > & f'(x_0) & > & \frac{f''(x_0)}{2!} & & \\
 & & & \vdots & & & & \\
n_1 \left\{\begin{array}{cc} x_0 & f_0 \\ x_1 & f_1 \\ x_1 & f_1 \\ x_1 & f_1 \\ \vdots & \vdots \end{array}\right. & & > & f[x_0,x_1] & > & \frac{f'(x_1)-f[x_0,x_1]}{x_1-x_0} & > & \dots \\
 & & > & f'(x_1) & > & \frac{f''(x_1)}{2!} & & \\
 & & > & f'(x_1) & & & & \\
 & & & \vdots & & & & \\
n_2 \left\{\begin{array}{cc} x_1 & f_1 \\ x_2 & f_2 \\ \vdots & \vdots \\ x_2 & f_1 \end{array}\right. & & > & f[x_2,x_1] & > & \frac{f'(x_2)-f[x_1,x_2]}{x_1-x_0} & > & \dots \\
 & & & \vdots & & & & \\
 & \vdots & & & & & &
\end{array}
$$

Nyní sestrojíme Newtonův interpolační polynom stejným způsobem jako dříve:

$$
\begin{aligned}
P_n(x) &= f(x_0)+(x-x_0)f'(x_0)+(x-x_0)^2\frac{f''(x_0)}{2}+\dots+ \\
&+ (x-x_0)^{n_0-1}\frac{f^{(n_0-1)}(x_0)}{(n_0-1)!}+(x-x_0)^{n_0}f[\underbrace{x_0,\dots,x_0}_{n_0},x_1]+ \\
&+ (x-x_0)^{n_0}(x-x_1)f[\underbrace{x_0,\dots,x_0}_{n_0},x_1,x_1]+\dots+ \\
&+ (x-x_0)^{n_0}(x-x_1)^{n_1-1}f[\underbrace{x_0,\dots,x_0}_{n_0},\underbrace{x_1,\dots,x_1}_{n_1}]+\dots+ \\
&+ (x-x_0)^{n_0}(x-x_1)^{n_1}\dots(x-x_m)^{n_m-1}f[\underbrace{x_0,\dots,x_0}_{n_0},\dots,\underbrace{x_m,\dots,x_m}_{n_m}].
\end{aligned}
$$

Lze ukázat, že tento polynom splňuje interpolační podmínky pro Hermitův polynom, t.j.

$$P^{(k)}(x_i) = f^{(k)}(x_i), \quad i = 0, \ldots, m, \quad k = 0, \ldots, n_i - 1$$

a je tedy totožný s hledaným Hermitovým interpolačním polynomem.

Poznámka 9. Pro výpočet hodnoty tohoto polynomu v daném bodě $\bar{x}$ lze také použít iterovanou interpolaci.

Příklad 6.12. Sestrojte Hermitův interpolační polynom, je-li dáno:

x_i	0	1	2
f_i	1	2	129
f_i'	0	7	448
f_i''	0	-	1344

Tabulka poměrných diferencí:

x_i	f_i						
0	$\underline{1}$						
		$\underline{0}$					
0	1		$\underline{0}$				
		0		$\underline{1}$			
0	1		1		$\underline{4}$		
		1		5		$\underline{11}$	
1	2		6		27		$\underline{6}$	
		7		57		23		$\underline{1}$
1	2		120		72		8	
		127		201		39		
2	129		321		150			
		448		351				
2	129		672					
		448						
2	129							

$$\begin{aligned} P_7(x) &= 1 + 0(x-0) + 0(x-0)^2 + 1(x-0)^3 + 4(x-0)^3(x-1) + \\ &+ 11(x-0)^3(x-1)^2 + 6(x-0)^3(x-1)^2(x-2) + \\ &+ 1(x-0)^3(x-1)^2(x-2)^2 = x^7 + 1\,. \end{aligned}$$

U Hermitovy interpolace se můžeme setkat s případy. kde v posloupnosti derivací zadaných v některém z uzlů jsou „mezery“ (lakunární interpolace). V těchto případech se může stát, že zadaná úloha nemá řešení, nebo má řešení více v závislosti na geometrické struktuře sítě uzlů interpolace a mezer v posloupnostech předepsaných hodnot. Zde uvedeme pouze ilustrační příklady, podrobněji se lze s problematikou lakunární interpolace seznámit například v přehledovém článku [10]. Jak lze snadno ověřit, úloha nalézt Hermitův interpolační polynom nemá řešení pro hodnoty:

x_i	x_0	x_1	x_2
f_i	f_0	$-$	f_2
f_i'		f_1'	

jestliže $x_1 = \frac{1}{2}(x_0 + x_2)$.
Naopak úloha pro hodnoty

x_i	x_0	x_1	x_2
f_i	f_0	$-$	f_2
f_i'		f_1'	
f_i''		f_1''	

má jediné řešení.

§ 6.9. Interpolace pomocí splajnů

Dosud uvedené interpolační metody aproximují danou funkci jedním interpolačním polynomem na celém intervalu. Tento postup není vždy výhodný, neboť lokální chování aproximované funkce ovlivňuje v tomto případě celkové chování aproximující funkce. Tato skutečnost vedla na myšlenku aproximace původní funkce analytickými funkcemi po částech. Takovými funkcemi jsou například *polynomiální splajny*. Jejich nejdůležitějším reprezentantem jsou *kubické* splajnové polynomy.[1]

Definice 6.6. Nechť je dána funkce f definovaná v intervalu $[a, b]$ a množina bodů, které nazýváme *uzly*, $a = x_0 < x_1 < \ldots < x_n = b$. Kubický interpolační splajn $S \in C^2[a, b]$ pro funkci f vyhovuje následujícím podmínkám:

a) S je kubickým polynomem S_j na subintervalu $[x_j, x_{j+1}]$ pro každé $j = 0, 1, \ldots, n-1$;

b) $S(x_j) = f(x_j)$, $j = 0, 1, \ldots, n$;

c) $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$, $j = 0, 1, \ldots, n-2$;

d) $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$, $j = 0, 1, \ldots, n-2$;

e) $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$, $j = 0, 1, \ldots, n-2$;

Jelikož při této konstrukci existují dva volné parametry, je možné požadovat, aby byly splněny jedny z následujících podmínek:

$$\begin{aligned} &\text{(i)} \quad S''(x_0) = S''(x_n) = 0, \\ &\text{(ii)} \quad S'(x_0) = f'(x_0),\ S'(x_n) = f'(x_n),\ (f \in C^1[a, b]) \end{aligned} \tag{6.43}$$

[1] Termín splajn je fonetickým přepisem anglického slova „spline“, které označuje zařízení na kreslení křivek. Jde o pružnou šablonu, která se vytvaruje do žádaného tvaru; v některých bodech se upevní závaží.

Splňuje-li kubický interpolační splajn S podmínku (i), nazývá se *přirozený* splajn, v případě podmínky (ii) jde o *úplný* splajn.

Poznámka 10. Je zřejmé, že podmínky b)–e) zaručují, že $S \in C^2[a,b]$.

Nyní uvedeme konstrukci kubického interpolačního splajnu S. Kubické polynomy na intervalech $[x_j, x_{j+1}]$, $j = 0, 1, \ldots, n-1$, uvažujme ve tvaru

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3$$

Je jasné, že

$$S_j(x_j) = a_j = f(x_j).$$

Z podmínky c) dále plyne

$$a_{j+1} = S_{j+1}(x_{j+1}) = S_j(x_{j+1}) =$$
$$= a_j + b_j(x_{j+1} - x_j) + c_j(x_{j+1} - x_j)^2 + d_j(x_{j+1} - x_j)^3, \ j = 0, 1, \ldots, n-2.$$

Zaveďme nyní označení

$$h_j = x_{j+1} - x_j, \qquad j = 0, 1, \ldots, n-1.$$

Dále položme $a_n = f(x_n)$. Z předchozího vztahu nyní plyne, že

$$a_{j+1} = a_j + b_j h_j + c_j h_j^2 + d_j h_j^3, \qquad j = 0, 1, \ldots, n-1. \tag{6.44}$$

Definujme obdobně $b_n = S'(x_n)$. Nyní

$$S_j'(x) = b_j + 2c_j(x - x_j) + 3d_j(x - x_j)^2$$

a odtud $S_j'(x_j) = b_j$, $j = 0, 1, \ldots, n-1$. Aplikací podmínky d) dostaneme

$$b_{j+1} = b_j + 2c_j h_j + 3d_j h_j^2, \qquad j = 0, 1, \ldots, n-1. \tag{6.45}$$

Položme nyní $c_n = S''(x_n)/2$ a aplikujme podmínku e). V tomto případě jsou výsledkem vztahy

$$c_{j+1} = c_j + 3d_j h_j, \qquad j = 0, 1, \ldots, n-1. \tag{6.46}$$

Nyní je naším úkolem určit koeficienty b_j, c_j, d_j, $j = 0, 1, \ldots, n$. Užitím vztahů (6.44), (6.45), (6.46) sestavíme systém rovnic pro neznámé koeficienty c_j. Nyní popíšeme tento postup. Z rovnice (6.46) vypočítáme d_j a dosadíme do rovnic (6.44) a (6.45) a dostaneme nové rovnice

$$a_{j+1} = a_j + b_j h_j + \frac{h_j^2}{3}(2c_j + c_{j+1}), \qquad j = 0, 1, \ldots, n-1 \tag{6.47}$$

a

$$b_{j+1} = b_j + h_j(c_j + c_{j+1}), \qquad j = 0, 1, \ldots, n-1. \tag{6.48}$$

Vyřešíme rovnici (6.47) nejdříve pro b_j:

$$b_j = \frac{1}{h_j}(a_{j+1} - a_j) - \frac{h_j}{3}(2c_j + c_{j+1}) \tag{6.49}$$

a pak zmenšíme index j o jedničku:

$$b_{j-1} = \frac{1}{h_{j-1}}(a_j - a_{j-1}) - \frac{h_{j-1}}{3}(2c_{j-1} + c_j) \tag{6.50}$$

Nyní dosadíme vyjádření (6.49) a (6.50) pro b_j a b_{j-1} do rovnice (6.48) (kde jsme snížili index o 1):

$$\begin{aligned} &h_{j-1}c_{j-1} + 2(h_{j-1} + h_j)c_j + h_j c_{j+1} = \\ &= \frac{3}{h_j}(a_{j+1} - a_j) - \frac{3}{h_{j-1}}(a_j - a_{j-1}),\ j = 1, 2, \ldots, n-1. \end{aligned} \tag{6.51}$$

Systém (6.51) je systém lineárních rovnic pro neznámé koeficienty c_j, $j = 0, \ldots, n$. Známe-li c_j, spočítáme ze vztahu (6.49) koeficienty b_j a ze vztahu (6.46) koeficienty d_j. Otázkou zůstává, zdali je soustava (6.51) řešitelná a jestliže ano, zda je řešení jediné. Odpověď na tuto otázku pro přirozené splajny dává následující věta.

Věta 6.13. *Nechť f je funkce definovaná na intervalu $[a, b]$. Pak f má jediný přirozený kubický interpolační splajn splňující podmínky $S''(a) = S''(b) = 0$.*

Důkaz. Nechť $\{x_i\}$, $i = 0, \ldots, n$, je dělení intervalu $[a, b]$: $a = x_0 < x_1 \ldots < x_n = b$. Okrajové podmínky (i) implikují, že

$$c_n = S''(x_n)/2 = 0, \qquad 0 = S''(x_0) = 2c_0 + 6d_0(x_0 - x_0),$$

tj. $c_0 = 0$. Rovnice $c_0 = 0$, $c_n = 0$ společně se systémem (6.51) tvoří lineární systém $A\boldsymbol{c} = \boldsymbol{g}$, $\boldsymbol{c} = (c_0, \ldots, c_n)^T$, kde $A \in \mathcal{M}_{n+1}$ a $\boldsymbol{b}$ je vektor dimenze $(n+1)$:

$$A = \begin{pmatrix} 1 & 0 & 0 & & \cdots & 0 \\ h_0 & 2(h_0+h_1) & h_1 & & \cdots & 0 \\ 0 & h_1 & 2(h_1+h_2) & h_2 & & \\ & & & \ddots & & \vdots \\ \vdots & \vdots & & h_{n-2} & 2(h_{n-2}+h_{n-1}) & h_{n-1} \\ 0 & 0 & & \cdots & & 1 \end{pmatrix},$$

$$\boldsymbol{g} = \begin{pmatrix} 0 \\ \frac{3}{h_1}(a_2 - a_1) - \frac{3}{h_0}(a_1 - a_0) \\ \vdots \\ \frac{3}{h_{n-1}}(a_n - a_{n-1}) - \frac{3}{h_{n-2}}(a_{n-1} - a_{n-2}) \\ 0 \end{pmatrix}.$$

Matice A je ryze řádkově diagonálně dominantní. Podle věty 4.2 je regulární a daná soustava má jediné řešení, tzn. existuje jediný kubický interpolační splajn. □

Obdobná věta platí i v případě, že jsou předepsány okrajové podmínky (ii).

Poznámka 11. Matice A je třídiagonální a pro řešení uvedeného systému lze užít Croutovy metody, neboť jsou splněny předpoklady věty 4.8.

Na závěr tohoto odstavce uvedeme odhad chyby pro okrajové podmínky (ii).

Věta 6.14. *Nechť $f \in C^4[a,b]$, $\max_{a\le x\le b}|f^{(4)}(x)| = M$. Pro kubický interpolační splajn S splňující okrajové podmínky $S'(a) = f'(a)$, $S'(b) = f'(b)$ platí*

$$\max_{a\le x\le b}|f(x) - S(x)| \le \frac{5M}{384}\max_{0\le j\le n-1}(x_{j+1} - x_j)^4.$$

Důkaz viz [4].

Odhad chyby pro přirozený splajn závisí rovněž na $(x_{j+1} - x_j)^4$, ale tento odhad lze velmi obtížně vyjádřit. Splajny hrají důležitou úlohu nejen při interpolaci funkcí, ale i při jiných typech aproximace. Lze je zkonstruovat tak, že zachovávají geometrický tvar funkce (např. konvexitu). Uplatňují se také ve statistice při vyhlazování dat.

Příklad 6.13. Sestrojte přirozený kubický interpolační splajn pro funkci $f(x) = 1/(1+x^2)$ na intervalu $[0,3]$. Za uzly zvolte body $x_0 = 0$, $x_1 = 1$, $x_2 = 3$.
Řešení. V tomto případě je třeba sestrojit 2 kubické polynomy S_0, S_1

$$S_0(x) = a_0 + b_0x + c_0x^2 + d_0x^3$$
$$S_1(x) = a_1 + b_1(x-1) + c_1(x-1)^2 + d_1(x-1)^3.$$

Je $h_0 = x_1 - x_0 = 1$, $h_1 = x_2 - x_1 = 2$, $a_0 = f(0) = 1$, $a_1 = f(1) = \frac{1}{2}$, $a_2 = f(3) = \frac{1}{10}$. Systém (6.51) je tvaru ($S''(x_0) = 0 = c_0$, $S''(x_2) = 0 = c_2$)

$$\begin{aligned} c_0 &= 0 \\ c_0h_0 + 2(h_0+h_1)c_1 + h_1c_2 &= \frac{3}{h_1}(a_2 - a_1) - \frac{3}{h_0}(a_1 - a_0) \\ c_2 &= 0. \end{aligned}$$

Odtud po dosazení za h_0, h_1, h_2, a_0, a_1, a_2 dostaneme

$$c_0 = 0, \qquad c_1 = \tfrac{3}{20}, \qquad c_2 = 0.$$

Dále užitím vztahů (6.49) resp. (6.50) vypočteme

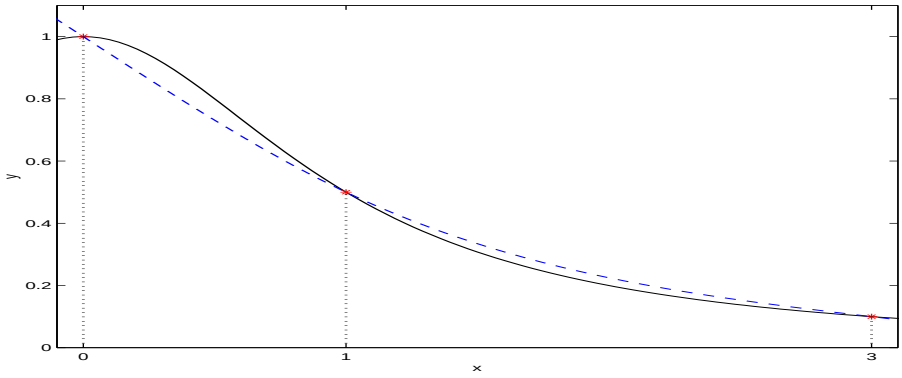
$$b_0 = -\tfrac{11}{20}, \qquad b_1 = -\tfrac{2}{5}.$$

A ze vztahů (6.46) vypočteme

$$d_0 = \tfrac{1}{20}, \qquad d_1 = -\tfrac{1}{40}.$$

Odpovídající kubické interpolační splajny jsou tvaru:

$$S_0(x) = 1 - \tfrac{11}{20}x + \tfrac{1}{20}x^3$$
$$S_1(x) = \tfrac{1}{2} - \tfrac{2}{5}(x-1) + \tfrac{3}{20}(x-1)^2 - \tfrac{1}{40}(x-1)^3.$$



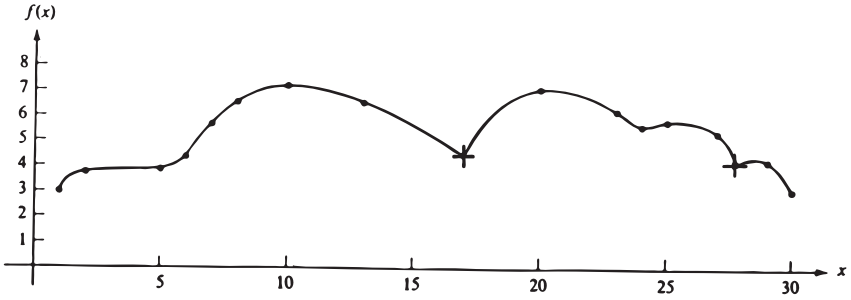
Obr. 6.11: Kubický splajn pro funkci $f(x) = 1/(1+x^2)$

Na závěr tohoto odstavce ukážeme použití splajnů v grafice (viz [4]). Pokusíme se v obrázku 6.12 aproximovat přirozenými kubickými splajny křivku, která tvoří jeho horní hranici (obrázek 6.13).

Obr. 6.12: Původní obrázek

Pomocí dostatečně jemné sítě je možné přibližně stanovit funkční hodnoty křivky, přičemž v místech, kde se křivka rychleji mění, je vhodné volit síť hustěji. Na dvou místech (označeny křížkem) je porušena hladkost křivky, proto ji rozdělíme na tři části a každou budeme aproximovat zvlášť.

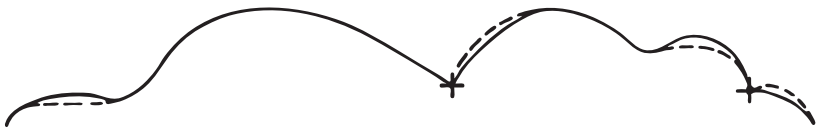
V tabulce 6.1 jsou uvedeny uzly a funkční hodnoty pro jednotlivé části křivky. Po výpočtu dostaneme tři splajny, jejich koeficienty jsou v tabulce 6.2. Obrázek 6.14 ukazuje rozdíl mezi původní křivkou (čárkovaně) a nalezenými splajny (plnou čarou).



Obr. 6.13: Aproximovaná křivka

Křivka 1		Křivka 2		Křivka 3	
x_i	$f(x_i)$	x_i	$f(x_i)$	x_i	$f(x_i)$
1	3,0	17	4,5	27,7	4,1
2	3,7	20	7,0	28	4,3
5	3,9	23	6,1	29	4,1
6	4,2	24	5,6	30	4,0
7	5,7	25	5,8		
8	6,6	27	5,2		
10	7,1	27,7	4,1		
13	6,7				
17	4,5				

Tabulka 6.1: Hodnoty bodů na jednotlivých křivkách



Obr. 6.14: Kubické interpolační splajny

Splajn 1

i	x_i	$a_i = f(x_i)$	b_i	c_i	d_i
0	1	3,0	0,786	0,000	-0,086
1	2	3,7	0,529	-0,257	0,034
2	5	3,9	-0,086	0,052	0,334
3	6	4,2	1,019	1,0583	-0,572
4	7	5,7	1,408	-0,664	0,156
5	8	6,6	0,547	-0,197	0,024
6	10	7,1	0,049	-0,052	-0,003
7	13	6,7	-0,342	-0,078	0,007
8	17	4,5			

Splajn 2

i	x_i	$a_i = f(x_i)$	b_i	c_i	d_i
0	17	4,5	1,106	0,000	-0,030
1	20	7,0	0,289	-0,272	0,025
2	23	6,1	-0,660	-0,044	0,204
3	24	5,6	-0,137	0,567	-0,230
4	25	5,8	0,306	-0,124	-0,089
5	27	5,2	-1,263	-0,660	0,314
6	27,7	4,1			

Splajn 3

i	x_i	$a_i = f(x_i)$	b_i	c_i	d_i
0	27,7	4,1	0,749	0,000	-0,910
1	28	4,3	0,503	-0,819	0,116
2	29	4,1	-0,787	-0,470	0,157
3	30	3,0			

Tabulka 6.2: Koeficienty jednotlivých splajnů

Cvičení ke kapitole 6

1. Najděte Lagrangeův interpolační polynom, je-li dáno

x_i	0	1	2	5
f_i	2	3	12	147

   ($P_3(x) = x^3 + x^2 - x + 2$.)

2. S jakou přesností lze vypočítat $\sqrt{115}$ pomocí Lagrangeova interpolačního polynomu pro funkci $y = \sqrt{x}$, když vybereme za uzly interpolace $x_0 = 100$, $x_1 = 121$, $x_2 = 144$?
   ($|E(115)| \leq 1{,}6.10^{-3}$.)

3. Pro případ ekvidistantních uzlů a tříbodového Lagrangeova vzorce najděte takový odhad veličiny $h^3 f'''(x)$, který v intervalu určeném třemi body zaručuje chybu metody menší než 10^{-d}, d je přirozené číslo. Použijte tohoto výsledku k odhadu největší hodnoty kroku h, kterého lze užít k interpolaci funkce $f(x) = \sin x$ na intervalu $[-\pi, \pi]$ s chybou menší než 10^{-10}.
   ($h \leq 1{,}15.10^{-3}$.)

4. Nechť l_i, $i = 0, 1, \ldots, n$ jsou fundamentální polynomy. Dokažte:

   a) Je-li $l_i(0) = c_i$, $i = 0, 1, \ldots, n$, pak

$$\sum_{i=0}^{n} c_i x_i^j = \begin{cases} 1 & \text{pro} \quad j = 0 \\ 0 & \text{pro} \quad j = 1, 2, \ldots, n \\ (-1)^n x_0 x_1 \ldots x_n & \text{pro} \quad j = n+1 \end{cases}$$

   (Návod: Využijte jednoznačnosti interpolačního polynomu.)

5. Nechť $x_0, \ldots, x_n$ jsou libovolná celá čísla $x_0 < x_1 < \ldots < x_n$. Ukažte, že každý algebraický polynom stupně n tvaru

$$Q(x) = x^n + a_1 x^{n-1} + \ldots + a_n$$

   nabývá v bodech $x_0, \ldots, x_n$ hodnot, z nichž alespoň jedna je v absolutní hodnotě větší nebo rovna $n!/2^n$.
   (Návod: Napište interpolační polynom pro Q v bodech $x_0, \ldots, x_n$, užijte jednoznačnosti a porovnejte koeficienty u x^n.)

6. Najděte Newtonův interpolační polynom, je-li dáno

x_i	0	2	3	5
f_i	1	3	2	5

   ($P_3(x) = \frac{3}{10}x^3 - \frac{13}{6}x^2 + \frac{62}{15}x + 1$.)

7. Je dána tabulka

x_i	-3	0	1	2
f_i	-13	2	3	12

Užitím inverzní interpolace najděte přibližně kořen rovnice $f(x) = 0$ ležící v intervalu $[-3, 0]$.
($\alpha \approx -2{,}13$.)

8. Vypočtěte fundamentální polynomy l_i, $i = 0, 1, \ldots, n$, jestliže za uzly interpolace zvolíme kořeny Čebyševova polynomu

$$T_{n+1}(x) = \cos\left((n+1)\arccos x\right).$$

($l_i(x) = \dfrac{\sqrt{1-x^2}(-1)^i \cos\left((n+1)\arccos x\right)}{(n+1)(x-x_i)}$, $x_i = \cos\dfrac{2i+1}{2(n+1)}\pi$, $i = 0, 1, \ldots, n$.)

9. a) Užijte vhodného Lagrangeova interpolačního polynomu stupně jedna, dva, tři a čtyři pro aproximaci hodnoty $f(2{,}5)$, jestliže je dáno

x_i	2,0	2,2	2,4	2,6	2,8
f_i	0,5103757	0,5207843	0,5104147	0,4813306	0,4359160

(

Uzly	stupeň	aproximace
2,4; 2,6	1	0,4958727
2,4; 2,6; 2,2	2	0,4982120
2,4; 2,6; 2,2; 2,8	3	0,4980630
všechny	4	0,4980705

)

   b) Navrhněte algoritmus pro obecný případ úlohy a).

10. Nechť $f(x) = \mathrm{e}^x$, $0 \le x \le 2$.

   a) Aproximujte $f(0{,}25)$ užitím lineární interpolace s uzly $x_0 = 0$, $x_1 = 0{,}5$.

   b) Aproximujte $f(0{,}75)$ užitím lineární interpolace s uzly $x_0 = 0{,}5$, $x_1 = 1$.

   c) Aproximujte $f(0{,}25)$ a $f(0{,}75)$ užitím kvadratické interpolace s uzly $x_0 = 0$, $x_1 = 1$, $x_2 = 2$.

   Které aproximace jsou lepší a proč?
   (a) 1,32436, b) 2,18350, c) 1,15277; 2,01191. Výsledky a), b) jsou lepší aproximací, neboť jsou zvoleny vhodnější uzly.)

11. Užijte Nevillova schematu pro určení aproximace ve cvičení 8.

12. a) Aproximujte $\sqrt{3}$ užitím Nevillova schematu pro funkci $f(x) = 3^x$ a uzly $x_0 = -2$, $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, $x_4 = 2$.
    b) Opakujte část a) užitím Aitkenova schematu.

13. Užitím iterované inverzní interpolace nalezněte přibližné řešení rovnice $x - \mathrm{e}^{-x} = 0$, je-li dáno

x_i	0,3	0,4	0,5	0,6
e^{x_i}	0,740818	0,670320	0,606531	0,548812

($f^{-1}(0) \approx 0{,}567142$.)

14. Aproximujte $f(0{,}05)$ užitím Newtonovy formule pro interpolaci vpřed, je-li dáno

x_i	0,0	0,2	0,4	0,6	0,8
$f(x_i)$	1,00000	1,22140	1,49182	1,82212	2,22554

($f(0{,}05) \approx 1{,}05126$.)

15. Jsou dány hodnoty funkce f: $f(a)$, $f(b)$, $f(c)$ v blízkosti jejího maxima nebo minima. Ukažte, že pro bod x, ve kterém se realizuje maximum nebo minimum, přibližně platí

$$x \approx \frac{(b^2 - c^2)f(a) + (c^2 - a^2)f(b) + (a^2 - b^2)f(c)}{2\{(b - c)f(a) + (c - a)f(b) + (a - b)f(c)\}}\,.$$

16. Sestrojte Hermitův interpolační polynom pro hodnoty

a)

x_i	-1	0	1
f_i	-1	0	1
f_i'	0	0	0

b)

x_i	0	1	2
f_i	1	-1	0
f_i'	0	0	
f_i''	0		

(a) $P_5(x) = \frac{1}{2}x^3(5 - 3x^2)$, b) $P_5(x) = (3x-1)(x-2)/2 + x(x-1)(x-2)(\frac{7}{4} + \frac{15}{4}x - \frac{33}{8}x^2)$.)

17. Užijte následujících hodnot pro konstrukci Hermitova interpolačního polynomu a pro určení hodnoty $\sin 0{,}34$.

x_i	$\sin x_i$	$(\sin x)'\vert_{x=x_i}$
0,30	0,29552	0,95534
0,32	0,31457	0,94924
0,33	0,32404	0,94604
0,35	0,34290	0,93937

($\sin 0{,}34 \approx 0{,}33350$.)

18. Interpolace funkce dvou proměnných.

    a) Lineární interpolace.
    Nechť $x_i \le \bar{x} \le x_{i+1}$, $y_i \le \bar{y} \le y_{i+1}$. Užitím lineární interpolace nejdříve pro x a pak pro y dokažte, že

    $$f(\bar{x},\bar{y}) \approx (1-\alpha)(1-\beta)f_{ij}+\beta(1-\alpha)f_{i,j+1}+\alpha(1-\beta)f_{i+1,j}+\alpha\beta f_{i+1,j+1},$$

    kde $\alpha = (\bar{x}-x_i)/(x_{i+1}-x_i)$, $\beta = (\bar{y}-y_j)/(y_{j+1}-y_j)$, $f_{ij} = f(x_i,y_j)$.

    b) Interpolace funkce dvou proměnných v obecném případě.
    Je dána tabulka

    $$\begin{array}{ccc} (a_1,b_1) & \cdots & (a_n,b_1) \\ (a_1,b_2) & \cdots & (a_n,b_2) \\ \vdots & & \vdots \\ (a_1,b_k) & \cdots & (a_n,b_k) \end{array}$$

    a hodnoty funkce $f(x,y)$ v těchto bodech. Užitím interpolace najděte přibližně hodnoty $f(\bar{x},\bar{y})$, $(\bar{x},\bar{y}) \neq (a_i,b_j)$, $i = 1,\ldots,n$, $j = 1,\ldots,k$.
    (Návod: Pro každý řádek tabulky sestrojte interpolační polynom a takto získaných hodnot užijte k interpolaci.)

19. Užitím cv. 18b řešte tuto úlohu. Je dána tabulka hodnot funkce $f(x,y) = \mathrm{e}^{x+y}$:

y \ x	0,7	0,9	1,5
0,7	4,05519	4,95303	9,0250
1,1	6,04964	7,38905	13,46373
1,3	7,38905	9,02501	16,44464

    Sestrojte interpolační polynom 2. stupně ve směru x a 2. stupně ve směru y a určete přibližnou hodnotu funkce v bodě $(1,1)$.
    ($f(1,1) \approx 7{,}38905$.)

20. Nalezněte přirozený kubický interpolační splajn pro $f(x) = \cos^2 x$ a uzly $x_0 = 0$, $x_1 = \frac{\pi}{2}$, $x_2 = \frac{3}{4}\pi$.
    ($S_0(x) = 1 - \frac{10}{3\pi}x + \frac{16}{3\pi^3}x^3$, $S_1(x) = \frac{2}{3\pi}(x-\frac{\pi}{2}) + \frac{8}{\pi^2}(x-\frac{\pi}{2})^2 - \frac{32}{3\pi^3}(x-\frac{\pi}{2})^3$.)

Kontrolní otázky ke kapitole 6

1. Je možné sestrojit Hermitův interpolační polynom pro následující hodnoty?

x_i	0	1	2
f_i'	1	0	1
f_i''	2		

2. lze pro danou množinu $(n+1)$ čísel c_0, c_1,...,c_n jediným způsobem sestrojit polynom $P \in \Pi_n$ splňující podmínky

$$P(x_0) = c_0, \quad P'(x_1) = c_1, \ \ldots, \ P^{(n)}(x_n) = c_n \, ?$$

3. Jsou dány dvojice čísel (x_i, f_i), $i = 0, 1, \ldots, n$, $x_i \neq x_k$ pro $i \neq k$.

   a) Lze najít právě jeden polynom Q stupně nejvýše $n-1$, který splňuje podmínky $Q(x_i) = f_i$, $i = 0, \ldots, n$?

   b) Lze najít právě jeden polynom R stupně alespoň $n+1$, který splňuje podmínky $R(x_i) = f_i$, $i = 0, \ldots, n$?

4. Je možné modifikovat Nevillovo, případně Aitkenovo schema na konstrukci Hermiteova polynomu?

5. Předpokládejme, že chceme řešit tuto úlohu interpolace funkce dvou proměnných:
   Nalezněte polynom P tvaru $P(x, y) = a_0 + a_1 x + a_2 y$ splňující podmínky $P(x_i, y_i) = f(x_i, y_i)$, $i = 0, 1, 2$.
   Je možné najít takový polynom pro libovolnou trojici bodů (x_i, y_i), $i = 0, 1, 2$?
   Jak to dopadne v případě, že $x^0 = (-1, -1)$, $x^1 = (0, 0)$, $x^2 = (1, 1)$?

6. V jakých případech bude splajn roven interpolačnímu polynomu?

Kapitola 7

Numerické derivování

Při řešení praktických úloh je někdy třeba najít derivaci funkce dané tabulkou. Může se také stát, že v důsledku složitého analytického vyjádření je bezprostřední výpočet derivace obtížný. V takových případech užíváme *numerického derivování*. Na základě poznatků z předchozí kapitoly je zřejmé, že formule pro numerické derivování lze získat derivací interpolačního polynomu a položit

$$f'(x) \approx P_n'(x).$$

Obecně však numerické derivování je operace méně přesná než interpolace, neboť ze skutečnosti, že hodnoty funkce a aproximujícího polynomu jsou blízké, neplyne ještě „blízkost“ hodnot derivací. Probereme nyní problém numerického derivování podrobněji.

§ 7.1. Numerický výpočet derivace

Úmluva. $I[x_0, \dots, x_n, x]$ bude označovat nejmenší uzavřený interval obsahující body $x_0, \dots, x_n, x$.

Jsou dány body (x_i, f_i), $i = 0, 1, \dots, n$, $x_i \neq x_k$ pro $i \neq k$. Nechť $P_n \in \Pi_n$ je Lagrangeův interpolační polynom pro tyto body, tj.

$$f(x) = P_n(x) + E(x), \tag{7.1}$$

kde

$$P_n(x) = \sum_{i=0}^{n} l_i(x) f_i, \qquad E(x) = \frac{\omega_{n+1}(x)}{(n+1)!} f^{(n+1)}(\xi), \qquad \xi = \xi(x),\ \xi \in I.$$

Derivujme vztah (7.1):

$$\begin{aligned} f'(x) &= P_n'(x) + E'(x) = \\ &= \sum_{i=0}^{n} l_i'(x) f_i + \frac{\omega_{n+1}'(x)}{(n+1)!} f^{(n+1)}(\xi) + \frac{\omega_{n+1}(x)}{(n+1)!} \frac{\mathrm{d}}{\mathrm{d}x} f^{(n+1)}(\xi) \end{aligned} \tag{7.2}$$

Vidíme, že chyba má v tomto případě složitější tvar než tomu bylo při interpolaci. Jestliže požadujeme výpočet derivace v některém z uzlových bodů x_j, což bývá nejčastější úloha, je předchozí formule tvaru

$$f'(x_j) = \sum_{i=0}^{n} l_i'(x_j) f_i + \frac{\omega_{n+1}'(x_j)}{(n+1)!} f^{(n+1)}(\xi_j), \qquad j = 0, 1, \ldots, n,\ \xi_j \in I.$$

Označme f_j' přibližnou hodnotu derivace v bodě x_j. Ta je tedy dána vztahem

$$f_j' = \sum_{i=0}^{n} l_i'(x_j) f_i$$

a výraz

$$\frac{\omega_{n+1}'(x_j)}{(n+1)!} f^{(n+1)}(\xi_j)$$

udává chybu této aproximace, $\xi_j = \xi_j(x_j)$. V obecném případě lze druhý člen chyby v (7.2) vyjádřit takto:

Věta 7.1. *Nechť*

$$\frac{\omega_{n+1}(x)}{(n+1)!} f^{(n+1)}(\xi), \quad \xi \in I[x_0, \ldots, x_n, x]$$

je chyba při Lagrangeově interpolaci. Nechť $f^{(n+2)}$ je spojitá v intervalu $I[x_0, \ldots, x_n, x]$. Pak existuje $\eta \in I[x_0, \ldots, x_n, x]$ takové, že

$$\frac{1}{(n+1)!} \frac{\mathrm{d}}{\mathrm{d}x} f^{(n+1)}(\xi) = \frac{1}{(n+2)!} f^{(n+2)}(\eta).$$

Důkaz. Uvažujme Lagrangeův interpolační polynom v bodě $x \neq x_i$, kde $i = 0, 1, \ldots, n$,

$$f(x) = \sum_{i=0}^{n} l_i(x) f_i + \frac{\omega_{n+1}(x)}{(n+1)!} f^{(n+1)}(\xi).$$

Pro l_i užijeme vyjádření

$$l_i(x) = \frac{\omega_{n+1}(x)}{(x - x_i)\, \omega_{n+1}'(x_i)},$$

a pak

$$f(x) = \sum_{i=0}^{n} \frac{\omega_{n+1}(x)}{(x - x_i)\, \omega_{n+1}'(x_i)} f_i + \frac{\omega_{n+1}(x)}{(n+1)!} f^{(n+1)}(\xi).$$

Protože $x \neq x_i$, $i = 0, 1, \ldots, n$, můžeme tuto rovnost vydělit $\omega_{n+1}(x)$ a derivovat

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{f(x)}{\omega_{n+1}(x)} \right) = -\sum_{i=0}^{n} \frac{f_i}{(x - x_i)^2 \omega_{n+1}'(x_i)} + \frac{1}{(n+1)!} \frac{\mathrm{d}}{\mathrm{d}x} f^{(n+1)}(\xi). \qquad (7.3)$$

Uvažujme ještě nyní další bod (x_{n+1}, f_{n+1}), $x_{n+1} \in I[x_0, \dots, x_n, x]$, $x_{n+1} \neq x, x_i$, $i = 0, 1, \dots, n$. Pak

$$\begin{aligned}
\omega_{n+2}(x) &= \prod_{i=0}^{n+1}(x - x_i) \quad \Rightarrow \quad \omega_{n+2}(x) = (x - x_{n+1})\,\omega_{n+1}(x) \\
\omega'_{n+2}(x) &= \omega_{n+1}(x) + (x - x_{n+1})\,\omega'_{n+1}(x) \quad \Rightarrow \\
&\Rightarrow \quad \omega'_{n+2}(x_i) = \begin{cases} (x_i - x_{n+1})\,\omega'_{n+1}(x_i) & \text{pro } i \neq n+1 \\ \omega_{n+1}(x_{n+1}) & \text{pro } i = n+1 \end{cases}
\end{aligned} \tag{7.4}$$

Sestrojme nyní Lagrangeův polynom pro body (x_i, f_i), $i = 0, 1, \dots, n+1$:

$$f(x) = \sum_{i=0}^{n+1} \frac{\omega_{n+2}(x)}{(x - x_i)\,\omega'_{n+2}(x_i)} f_i + \frac{\omega_{n+2}(x)}{(n+2)!} f^{(n+2)}(\tau), \tag{7.5}$$

$\tau = \tau(x) \in I[x_0, \dots, x_{n+1}, x]$. Vztah (7.5) nyní poněkud upravíme:

$$\begin{aligned}
&f(x) - \frac{\omega_{n+2}(x)}{(x - x_{n+1})\,\omega'_{n+2}(x_{n+1})} f_{n+1} = \\
&= \sum_{i=0}^{n} \frac{\omega_{n+2}(x)}{(x - x_i)\,\omega'_{n+2}(x_i)} f_i + \frac{\omega_{n+2}(x)}{(n+2)!} f^{(n+2)}(\tau).
\end{aligned}$$

V dalším použijeme vztahů (7.4):

$$\begin{aligned}
&f(x) - \frac{\omega_{n+1}(x)}{\omega_{n+1}(x_{n+1})} f_{n+1} = \\
&= \sum_{i=0}^{n} \frac{(x - x_{n+1})\,\omega_{n+1}(x)}{(x - x_i)(x_i - x_{n+1})\,\omega'_{n+1}(x_i)} f_i + \frac{\omega_{n+2}(x)}{(n+2)!} f^{(n+2)}(\tau).
\end{aligned}$$

Tuto rovnici vydělíme $(x - x_{n+1})\,\omega_{n+1}(x)$:

$$\begin{aligned}
&\frac{f(x)}{(x - x_{n+1})\,\omega_{n+1}(x)} - \frac{f_{n+1}}{(x - x_{n+1})\,\omega_{n+1}(x_{n+1})} = \\
&= \sum_{i=0}^{n} \frac{f_i}{(x - x_i)(x_i - x_{n+1})\,\omega'_{n+1}(x_i)} + \frac{1}{(n+2)!} f^{(n+2)}(\tau)
\end{aligned}$$

neboli

$$\begin{aligned}
&\frac{\dfrac{f(x)}{\omega_{n+1}(x)} - \dfrac{f_{n+1}}{\omega_{n+1}(x_{n+1})}}{x - x_{n+1}} = \\
&= \sum_{i=0}^{n} \frac{f_i}{(x - x_i)(x_i - x_{n+1})\,\omega'_{n+1}(x)} + \frac{1}{(n+2)!} f^{(n+2)}(\tau).
\end{aligned}$$

Přechodem k limitě pro $x_{n+1} \to x$ dostaneme ($f^{(n+2)}$ spojitá v I)

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{f(x)}{\omega_{n+1}(x)} = -\sum_{i=0}^{n}\frac{f_i}{(x-x_i)^2\omega'_{n+1}(x_i)} + \frac{1}{(n+2)!}f^{(n+2)}(\eta), \qquad (7.6)$$
$$\eta \in I[x_0,\ldots,x_n,x].$$

Porovnáním (7.5) a (7.6) plyne tvrzení věty. □

Výsledná formule pro výpočet derivace v bodě $x \neq x_i$, $i = 0,1,\ldots,n$, je tvaru

$$f'(x) = \sum_{i=0}^{n} l'_i(x)f_i + \frac{\omega'_{n+1}(x)}{(n+1)!}f^{(n+1)}(\xi) + \frac{\omega_{n+1}(x)}{(n+2)!}f^{(n+2)}(\eta),$$
$$\xi,\eta \in I[x_0,\ldots,x_n,x]$$

Poznámka 1. Pro vyšší derivace platí obdobný vztah

$$\frac{1}{n!}\frac{\mathrm{d}^k}{\mathrm{d}x^k}f^{(n)}(\xi) = \frac{k!}{(n+k)!}f^{(n+k)}(\eta).$$

V praxi se často setkáváme s případem, kdy množina uzlů x_i, $i = 0,1,\ldots,n$ je ekvidistantní. Je-li celkový počet uzlů lichý, je vhodné přiřadit uzlům kladné a záporné indexy a to takto:

$$x_{-l},\ldots,x_{-1},x_0,x_1,\ldots,x_l,$$
$$x_i = x_0 + ih, \qquad i = \pm1,\ldots,\pm l,\ h > 0.$$

Příklad 7.1. Odvoďte formuli pro výpočet derivace v prostředním ze tří uzlů: x_{-1}, x_0, x_1, $x_i = x_0 + ih$, $i = \pm1$.
Řešení. Sestrojíme Lagrangeův interpolační polynom pro hodnoty

x_i	x_{-1}	x_0	x_1
f_i	f_{-1}	f_0	f_1

Je $f(x) = P_2(x) + E(x)$

$$f(x) = \frac{(x-x_0)(x-x_1)}{(x_{-1}-x_0)(x_{-1}-x_1)}f_{-1} + \frac{(x-x_{-1})(x-x_1)}{(x_0-x_{-1})(x_0-x_1)}f_0 +$$
$$+ \frac{(x-x_{-1})(x-x_0)}{(x_1-x_{-1})(x_1-x_0)}f_1 + \frac{\omega_3(x)}{3!}f'''(\xi),$$

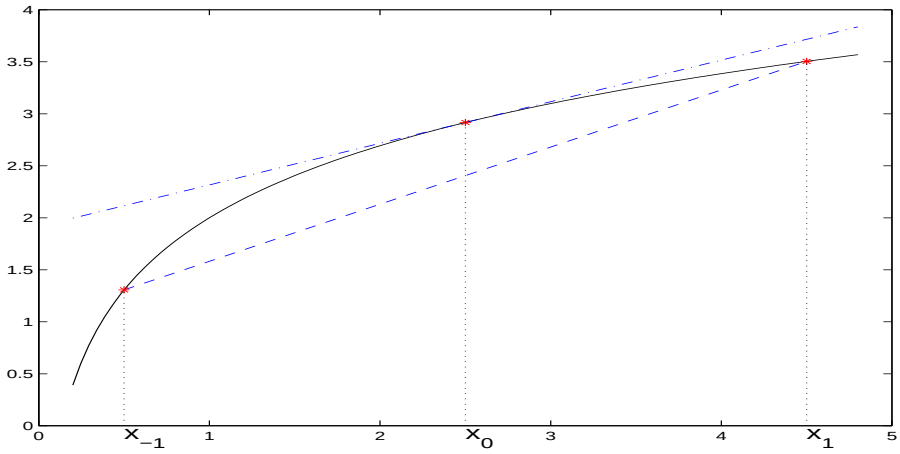
$\omega_3(x) = (x-x_{-1})(x-x_0)(x-x_1)$. Derivujeme

$$f'(x) = \frac{2x-x_0-x_1}{2h^2}f_{-1} - \frac{2x-x_{-1}-x_1}{h^2}f_0 + \frac{2x-x_{-1}-x_0}{2h^2}f_1 +$$
$$+ \frac{1}{3!}\omega'_3(x)f'''(\xi) + \frac{1}{3!}\omega_3(x)\frac{\mathrm{d}}{\mathrm{d}x}(f'''(\xi)),$$

kde jsme položili $x_i = x_0 + ih$, $i = \pm 1$, a pro $x = x_0$ máme

$$f'(x_0) = \frac{1}{2h}(-f_{-1} + f_1) - \frac{h^2}{6} f'''(\xi_0). \tag{7.7}$$

Všimněme si ještě geometrického významu této formule — viz obr. 7.1. Zde $f'_0 = (-f_{-1} + f_1)/2h$ je přibližná hodnota derivace v bodě x_0, což geometricky znamená směrnici sečny určené body (x_{-1}, f_{-1}) a (x_1, f_1). Stejným způsobem lze odvodit



Obr. 7.1: Sečna určená body (x_{-1}, f_{-1}), (x_1, f_1) (čárkovaně) a tečna v bodě (x, f_0) (čerchovaně)

i formule pro výpočet derivace v bodech x_{-1}, x_1:

$$f'(x_{-1}) = \frac{1}{2h}(-3f_{-1} + 4f_0 - f_1) + \frac{h^2}{3} f'''(\xi_{-1}) \tag{7.8}$$

$$f'(x_1) = \frac{1}{2h}(f_{-1} - 4f_0 + 3f_1) + \frac{h^2}{3} f'''(\xi_1) \tag{7.9}$$

Tyto formule se nazývají *tříbodové*.

Příklad 7.2. Užitím formulí (7.7), (7.8) a (7.9) vypočtěte derivaci funkce $f(x) = \log x + 2$ v bodech 0,5, 1, 1,5; $h = 0,5$.
Řešení:

$$f'_{-1} = \frac{1}{2h}(-3\log(0{,}5) + 4\log(1) - \log(1{,}5)) = 1{,}6740,$$
$$f'_0 = \frac{1}{2h}(-\log(0{,}5) + \log(1{,}5)) = 1{,}0986,$$
$$f'_1 = \frac{1}{2h}(\log(0{,}5) - 4\log(1) + 3\log(1{,}5)) = 0{,}5232.$$

Přesné hodnoty jsou $f'(0{,}5) = 2$, $f'(1) = 1$, $f'(1{,}5) = 2/3$.

Poznámka 2. Všimněme si, že chyba ve vzorci (7.7) je rovna přibližně polovině chyby ve vzorcích (7.8), (7.9). To je logické, neboť (7.7) užívá hodnot v bodech ležících po obou stranách x_0, ale vzorce (7.8), (7.9) pouze hodnot ležících pouze na jednu stranu od x_{-1} resp. x_1. Formule (7.8) a (7.9) jsou tedy vhodné pro výpočet derivace v blízkosti koncových bodů intervalu, neboť nemusíme mít k dispozici hodnoty vně intervalu.

Při použití formulí pro numerické derivování se setkáváme ještě s dalším problémem. Jestliže hodnoty f_i jsou dány s chybou ε_i, může tato okolnost podstatně ovlivnit výslednou hodnotu f_i'. Ukážeme to na případě formule (7.7). Nechť f_i je přesná hodnota, $\tilde{f}_i$ je přibližná hodnota v bodě x_i, $i = 0, \pm 1$. Celkovou chybu T lze odhadnout takto:

$$|T| \leq \frac{1}{2h}(|\tilde{f}_{-1} - f_{-1}| + |\tilde{f}_1 - f_1|) + \frac{h^2}{6}\,|f'''(\xi_0)| \leq \frac{1}{2h}(\varepsilon_{-1} + \varepsilon_1) + \frac{h^2}{6}\,|f'''(\xi_0)|\,,$$

kde $|f_i - \tilde{f}_i| \leq \varepsilon_i$; položme

$$\varepsilon = \max(\varepsilon_1, \varepsilon_{-1}), \qquad M_3 = \max_{[x_{-1}, x_1]} |f'''(x)|\,.$$

Pak

$$|T| \leq \frac{\varepsilon}{h} + \frac{h^2}{6} M_3.$$

První člen chyby (např. chyba způsobená zaokrouhlováním) závisí nepřímo úměrně na h, druhý člen (chyba metody) závisí přímo úměrně na h. Vzniká problém, jak volit h, aby celková chyba byla minimální. Hledejme tedy minimum funkce

$$g(h) = \frac{\varepsilon}{h} + \frac{h^2}{6} M_3.$$

Ze vztahu $g'(h) = 0$ dostáváme bod minima

$$h_{\text{opt}} = \sqrt[3]{\frac{3\varepsilon}{M_3}}.$$

Pro obecný případ je problém podrobně popsán v [1], [19].

Poznámka 3. V případě, že hodnoty f_i jsou dány s malou přesností (např. byly získány empiricky), není vhodné použít formuli pro numerické derivování přímo, neboť by mohlo dojít ke zkreslení výsledků. V takových případech je lépe nejdříve naměřené hodnoty „vyrovnat“ metodou nejmenších čtverců a pak teprve derivovat.

Postupem uvedeným výše lze získat formule pro numerický výpočet derivací vyšších řádů. Jako příklady lze uvést tyto tříbodové formule:

$$\begin{aligned}
f''(x_{-1}) &= \frac{1}{h^2}(f_{-1} - 2f_0 + f_1) - h f'''(\xi_{-1}) + \frac{h^2}{6} f^{(4)}(\eta_{-1}) \\
f''(x_0) &= \frac{1}{h^2}(f_{-1} - 2f_0 + f_1) - \frac{h^2}{12} f^{(4)}(\eta_0) \\
f''(x_1) &= \frac{1}{h^2}(f_{-1} - 2f_0 + f_1) + h f'''(\xi_1) - \frac{h^2}{6} f^{(4)}(\eta_1)\,.
\end{aligned}$$

§ 7.2. Diferenční aproximace

Zmíníme se ještě o jiném způsobu přibližného výpočtu derivace. Při numerickém řešení diferenciálních rovnic se u některých metod aproximují derivace diferencemi (diferenční metody). Ukážeme, jak se v takových případech aproximují první a druhé derivace. Stejně jako v příkladě 7.1 uvažujme tři body $x_{-1} = x_0 - h$, x_0, $x_1 = x_0 + h$. Předpokládejme, že f má dostatečný počet derivací v okolí x_0. Napišme nyní Taylorův rozvoj v bodě x_0:

$$\begin{aligned} f(x_0+h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \frac{h^3}{3!}f^{(3)}(x_0) + \\ &\quad + \frac{h^4}{4!}f^{(4)}(x_0) + \frac{h^5}{5!}f^{(5)}(x_0) + \frac{h^6}{6!}f^{(6)}(x_0) + O(h^7), \end{aligned} \tag{7.10}$$

$$\begin{aligned} f(x_0-h) &= f(x_0) - hf'(x_0) + \frac{h^2}{2}f''(x_0) - \frac{h^3}{3!}f^{(3)}(x_0) + \\ &\quad + \frac{h^4}{4!}f^{(4)}(x_0) - \frac{h^5}{5!}f^{(5)}(x_0) + \frac{h^6}{6!}f^{(6)}(x_0) + O(h^7)\,. \end{aligned} \tag{7.11}$$

Ze vztahu (7.10) nyní dostaneme aproximaci první derivace ve tvaru

$$f'(x_0) = \frac{f(x_0+h) - f(x_0)}{h} + O(h)\,.$$

Výraz $(f(x_0+h) - f(x_0))/h$ se nazývá *pravá diferenční derivace*.

Ze vztahu (7.10) plyne

$$f'(x_0) = \frac{f(x_0) - f(x_0-h)}{h} + O(h)\,.$$

Výraz $(f(x_0) - f(x_0-h))/h$ se nazývá *levá diferenční derivace*.

Pravá a levá diferenční derivace aproximují $f'(x_0)$ s chybou řádu $O(h)$ (tj. chyba se chová přibližně jako kh, $k = konst.$). Jestliže vztahy (7.10) a (7.11) odečteme, dostaneme aproximaci derivace, která je řádu $O(h^2)$:

$$f'(x_0) = \frac{f(x_0+h) - f(x_0-h)}{2h} + O(h^2)\,.$$

Tato aproximace se nazývá *centrální diferenční derivace*. Součtem vztahů (7.10) a (7.11) se získá aproximace druhé derivace v bodě x_0:

$$f''(x_0) = \frac{f(x_0+h) - 2f(x_0) + f(x_0-h)}{h^2} + O(h^2)\,.$$

Poznámka 4. Všimněme si, že ve všech uvedených příkladech jsme neznámé hodnoty $f'(x_0)$, $f''(x_0)$ aproximovali jistými formulemi závislými na kroku h, který zde hraje roli parametru. V následujícím odstavci ukážeme, jak lze těchto formulí užít k získání aproximací vyšších řádů.

§ 7.3. Richardsonova extrapolace

Technika známá jako *Richardsonova extrapolace* se často používá pro získání výsledků vyšších řádů přesnosti užitím formulí nižších řádů přesnosti. Myšlenka této metody pochází už od Archimeda (cca 200 př. n. l.). Vysvětlíme nejdříve obecný postup a pak ukážeme aplikaci na numerický výpočet derivace.

Nechť $N(h)$ je formule, která aproximuje neznámou veličinu M a nechť řád této aproximace je $O(h^2)$. Navíc předpokládejme, že aproximace $N(h)$ veličiny M může být vyjádřena ve tvaru:

$$M = N(h) + k_1 h^2 + O(h^4)\,, \tag{7.12}$$

kde k_1 je konstanta. Napíšeme-li tento vztah s $h/2$ místo s h, dostaneme

$$M = N(\frac{h}{2}) + k_1(\frac{h}{2})^2 + O((\frac{h}{2})^4)\,. \tag{7.13}$$

Vynásobme rovnici (7.13) čtyřmi a odečteme od ní rovnici (7.12):

$$3M = 4N(\frac{h}{2}) - N(h) + O(h^4)\,,$$

odtud

$$M = \frac{4N(\frac{h}{2}) - N(h)}{3} + O(h^4)\,.$$

Položme $N_1(h) = N(h)$ a

$$N_2(h) = \frac{4N_1(\frac{h}{2}) - N_1(h)}{3}. \tag{7.14}$$

Veličina $N_2(h)$ je novou aproximací veličiny M a řád této aproximace je $O(h^4)$. Tento postup lze zobecnit takto: Předpokládejme, že chyba aproximace $N(h)$ veličiny M může být vyjádřena ve tvaru

$$M = N(h) + \sum_{j=1}^{m-1} k_j h^{2j} + O(h^{2m})\,,$$

kde $k_1,\ldots,k_{m-1}$ jsou konstanty. Aproximace $N_j(h)$ řádu $O(h^{2j})$, $j = 2, 3, \ldots, m$ jsou definovány vztahy

$$N_j(h) = \frac{4^{j-1}N_{j-1}(\frac{h}{2}) - N_{j-1}(h)}{4^{j-1} - 1} \tag{7.15}$$

a výpočet lze uspořádat do tabulky:

$$\begin{array}{ccccc} N_1(h) & & & & \\ N_1(\frac{h}{2}) & N_2(h) & & & \\ N_1(\frac{h}{4}) & N_2(\frac{h}{2}) & N_3(h) & & \\ N_1(\frac{h}{8}) & N_2(\frac{h}{4}) & N_3(\frac{h}{2}) & N_4(h) & \\ N_1(\frac{h}{16}) & N_2(\frac{h}{8}) & N_3(\frac{h}{4}) & N_4(\frac{h}{2}) & N_5(h) \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

Poznámka 5. Uvedený postup se nazývá *Richardsonova extrapolace*. Snažíme se totiž získat hodnotu pro $h \to 0$, tj. jedná se o extrapolaci z kladných hodnot h.

Je zřejmé, že centrální diferenční derivace může být vyjádřena ve tvaru (viz (7.10) a (7.10))

$$f'(x_0) = \frac{f(x_0+h) - f(x_0-h)}{2h} - \frac{h^2}{6} f'''(x_0) - \frac{h^4}{120} f^{(5)}(x_0) + O(h^6)\,. \qquad (7.16)$$

V tomto případě je

$$N_1(h) = \frac{f(x_0+h) - f(x_0-h)}{2h}, \quad k_1 = -\frac{f'''(x_0)}{6}, \quad k_2 = -\frac{f^{(5)}(x_0)}{120}\,. \qquad (7.17)$$

Příklad 7.3. Užijte formuli (7.16) a (7.15) pro výpočet druhé derivace funkce $f(x) = x\,e^x$ v bodě $x_0 = 2$ s krokem $h = 0,2$
Řešení: Je

$$N_1(h) = N_1(0,2) = \frac{1}{0,4}(f(2,2) - f(1,8)) = 22,414160$$

$$\begin{aligned} N_1(0,1) &= 22,228787 \\ N_1(0,05) &= 22,182565 \end{aligned}$$

Další aproximace jsou uspořádány v tabulce

$$\begin{array}{lll} N_1(0,2) = 22,414160 & & \\ N_1(0,1) = 22,228787 & N_2(0,2) = \frac{4\,N_1(0,1) - N_1(0,2)}{3} = 22,166996 & \\ N_1(0,05) = 22,182565 & N_2(0,1) = \frac{4\,N_1(0,05) - N_1(0,1)}{3} = 22,167158 & \\ & & N_3(0,2) = \frac{16\,N_2(0,1) - N_2(0,2)}{15} = 22,167168\,. \end{array}$$

Pro srovnání – přesná hodnota $f'(x) = e^x + x\,e^x$ v bodě $x_0 = 2$ je $22,167168$.

Poznámka 6. Jak uvidíme v kapitole 9, lze Richardsonovy extrapolace s výhodou použít i pro numerický výpočet integrálu.

Cvičení ke kapitole 7

1. Odvoďte formuli (7.8), (7.9).

2. Odvoďte pětibodovou formuli ve tvaru

$$f_0' = \frac{1}{12h}(f_{-2} - 8f_{-1} + 8f_1 - f_2) + \frac{h^4}{30}f^{(5)}(\xi), \qquad x_{-2} < \xi < x_2.$$

3. Užitím formulí (7.7), (7.8), (7.9) vypočtěte derivace funkce v daných bodech

x_i	$-0{,}3$	$-0{,}1$	$0{,}1$	$0{,}3$
f_i	$-0{,}20431$	$-0{,}08993$	$0{,}11007$	$0{,}39569$

($f'(-0{,}3) \approx 0{,}35785$, $f'(-0{,}1) \approx 0{,}78595$, $f'(0{,}1) \approx 1{,}2141$, $f'(0{,}3) \approx 61{,}6422$.)

4. a) Nechť $f(x) = 2^x \sin x$. Aproximujte hodnotu $f'(1{,}05)$ užitím $h = 0{,}05$ a $h = 0{,}01$ ve formuli (7.7), jsou-li dány hodnoty:

x_i	1,0	1,04	1,06	1,10
$f(x_i)$	1,6829420	1,7732994	1,8188014	1,9103448

   b) Opakujte část a) pro případ, že všechny funkční hodnoty zaokrouhlíte na čtyři desetinná místa.
   (2,27403, 2,27510.)

5. Užitím formule (7.7) najděte první derivaci funkce $f(x) = 1/(1+x)$ v bodě $x = 0{,}005$. Užijte a) $h = 1{,}0$, b) $h = 0{,}01$ a výsledky porovnejte s přesnou hodnotou. Vysvětlete!

6. Použijte formule (7.16) a (7.15) k výpočtu hodnoty $N_3(h)$ pro následující funkce a kroky h:

   a) $f(x) = \ln x$, $x_0 = 1,0$, $h = 0,4$

   b) $f(x) = x + e^x$, $x_0 = 0,0$, $h = 0,4$

   c) $f(x) = 2^x \sin x$, $x_0 = 1,05$, $h = 0,4$

   Výsledek porovnejte s přesnými hodnotami.

Kontrolní otázky ke kapitole 7

1. Může být první krok Richardsonovy extrapolace, tj. vztah (7.14), popsán pomocí interpolace určené body $(h^2, N_1(h))$, $((\frac{h}{2})^2, N_1(\frac{h}{2}))$? Jaká je hodnota příslušného interpolačního polynomu v bodě 0?

2. Je výhodné použít pro numerický výpočet derivace podle formule (7.2), ve které jsou jako uzly použity kořeny Čebyševova polynomu (viz kapitola 6)? (Návod: $|T_n'(x)| \leq n^2$ pro $x \in [-1, 1]$.)

3. Je možné použít hodnot z příkladu 1 pro výpočet $f'''(x_i)$, $i = -1,\ 0,\ 1$? Stačí v tomto případě aproximace funkce polynomem druhého stupně?

Kapitola 8

Ortogonální polynomy

V této kapitole se budeme zabývat určitými polynomy, které budou velmi užitečné při konstrukci formulí numerického integrování. Zde uvedeme pouze definice a nejdůležitější vlastnosti. Podrobným studiem těchto polynomů se zabývá např. publikace [22].

Nechť Π_j stejně jako dříve značí množinu všech polynomů stupně nejvýše j a $\bar{\Pi}_j$ množinu všech normovaných polynomů stupně j, tj. polynomů s koeficientem rovným jedné u nejvyšší mocniny.

Nechť w je funkce, o které předpokládáme, že je integrovatelná a nezáporná na intervalu $[a, b]$ a $w(x) > 0$ skoro všude na $[a, b]$. Takovou funkci budeme nazývat *vahovou funkcí*.

Dále definujeme skalární součin

$$\langle f, g \rangle = \int_a^b w(x) f(x) g(x) \, \mathrm{d}x$$

pro všechny funkce, pro které existuje konečný integrál

$$\langle f, f \rangle = \int_a^b w(x) f^2(x) \, \mathrm{d}x < +\infty.$$

Jestliže $\langle f, g \rangle = 0$, říkáme, že funkce f, g jsou *ortogonální na intervalu* $[a, b]$ *s vahou* w.

Následující věta dokazuje existenci posloupnosti navzájem ortogonálních polynomů vzhledem k vahové funkci w.

Věta 8.1. *Pro danou vahovou funkci w na $[a, b]$ existují polynomy $p_j \in \bar{\Pi}_j$, $j = 0, 1, 2, \ldots$, takové, že*

$$\langle p_i, p_k \rangle = 0 \quad \textit{pro} \quad i \neq k. \tag{8.1}$$

Tyto polynomy jsou jednoznačně definovány vztahy

$$p_0(x) \equiv 1$$

$$p_{i+1}(x) = (x - \delta_{i+1})p_i(x) - \gamma_i^2 p_{i-1}(x) \quad pro \quad i \geq 0,$$

kde $p_{-1}(x) \equiv 0$ *a*

$$\delta_{i+1} = \langle xp_i, p_i\rangle / \langle p_i, p_i\rangle \quad pro \quad i \geq 0$$

$$\langle xp_i, p_i\rangle = \int_a^b w(x) x p_i^2(x)\,\mathrm{d}x$$

a

$$\gamma_i^2 = \begin{cases} 0 & pro \quad i = 0, \\ \langle p_i, p_i\rangle / \langle p_{i-1}, p_{i-1}\rangle & pro \quad i \geq 1. \end{cases}$$

Důkaz. Polynomy p_j, $j = 0, 1, 2, \ldots$, lze sestrojit rekurentně pomocí Gramova-Schmidtova ortogonalizačního procesu ([22]). □

Každý polynom $p \in \Pi_k$ lze zřejmě vyjádřit jako lineární kombinaci ortogonálních polynomů $p_i \in \bar{\Pi}_i$, $i \leq k$. Máme tedy:

Důsledek. $\langle p, p_n\rangle = 0$ *pro všechny polynomy* $p \in \Pi_{n-1}$, $p_n \in \bar{\Pi}_n$.

Věta 8.2. *Nechť* $\{p_j\}$ *je systém polynomů ortogonálních s vahou* w *na intervalu* $[a, b]$.

Platí: Každý polynom p_j *má všechny kořeny reálné, různé a všechny leží v intervalu* (a, b).

Důkaz lze najít v [22].

Některé vahové funkce se vyskytují v praxi dosti často, příslušné ortogonální polynomy se uvažují ve standardním tvaru, v němž je koeficient u x^n obvykle různý od jedné. Uvedeme nyní některé speciální ortogonální polynomy.

1. *Legendrovy polynomy* P_n jsou ortogonální na intervalu $[-1, 1]$ s vahou $w(x) \equiv 1$.

   Vlastnosti:

   a) Ortogonalita

   $$\int_{-1}^{1} P_n(x)P_m(x)\,\mathrm{d}x = \begin{cases} 0 & \text{pro} \quad n \neq m, \\ \dfrac{2}{2n+1} & \text{pro} \quad n = m. \end{cases} \tag{8.2}$$

   b) Platí rekurentní vztah

   $$P_{n+1}(x) = \frac{2n+1}{n+1} x P_n(x) - \frac{n}{n+1} P_{n-1}(x), \qquad n = 1, 2, 3, \ldots$$

   Je

   $$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, \ldots$$

   c) Polynomy P_n vyhovují diferenciální rovnici

   $$(1 - x^2)y'' - 2xy' + n(n+1)y = 0.$$

2. *Čebyševovy polynomy* T_n jsou ortogonální na intervalu $[-1,1]$ s vahou $w(x) = 1/\sqrt{1-x^2}$.

   Vlastnosti:

   a) Ortogonalita

$$\int_{-1}^{1} \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}\,\mathrm{d}x = \begin{cases} 0 & \text{pro} \quad n \neq m, \\ \pi & \text{pro} \quad n = m = 0, \\ \frac{\pi}{2} & \text{pro} \quad n = m \neq 0. \end{cases} \tag{8.3}$$

   b) Platí rekurentní vztah

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

   Je

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \ldots$$

   c) Polynomy T_n jsou řešením diferenciální rovnice

$$(1-x^2)y'' - xy' + n^2 y = 0.$$

   S těmito polynomy jsme se již setkali, když jsme hledali polynomy s nejmenší odchylkou od nuly na intervalu $[-1,1]$.

3. *Laguerrovy polynomy* L_n jsou ortogonální na intervalu $[0,\infty)$ s vahou $w(x) = x^\alpha \mathrm{e}^{-x}$, $\alpha > -1$.

   Vlastnosti:

   a) Ortogonalita

$$\int_0^\infty x^\alpha \mathrm{e}^{-x} L_n(x,\alpha)L_m(x,\alpha)\,\mathrm{d}x = \begin{cases} 0 & \text{pro} \quad n \neq m, \\ \dfrac{\Gamma(\alpha+n+1)}{n!} & \text{pro} \quad n = m, \end{cases} \tag{8.4}$$

   kde Γ je gamma funkce.

   b) Platí rekurentní vztah

$$L_{n+1}(x,\alpha) = \frac{2n+\alpha+1-x}{n+1}L_n(x,\alpha) - \frac{n+\alpha}{n+1}L_{n-1}(x,\alpha)$$

   Je

$$L_0(x,\alpha) = 1, \quad L_1(x,\alpha) = -x + \alpha + 1, \ \ldots$$

   c) Polynomy L_n jsou řešením diferenciální rovnice

$$xy'' + (\alpha - x + 1)y' + ny = 0.$$

4. *Hermitovy polynomy* H_n jsou ortogonální na intervalu $(-\infty, +\infty)$ s vahou $w(x) = \mathrm{e}^{-x^2}$.

   Vlastnosti:

   a) Ortogonalita

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2} H_n(x) H_m(x)\,\mathrm{d}x = \begin{cases} 0 & \text{pro} \quad n \neq m, \\ 2^n n! \sqrt{\pi} & \text{pro} \quad n = m. \end{cases} \tag{8.5}$$

   b) Platí rekurentní vztah

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$$

   Je

$$H_0(x) = 1, \quad H_1(x) = 2x, \quad H_2(x) = 4x^2 - 2, \ldots$$

   c) Polynomy H_n jsou řešením diferenciální rovnice

$$y'' - 2xy' + 2ny = 0.$$

Další vlastnosti ortogonálních polynomů p_n dává následující věta.

Věta 8.3. *Nechť $\{p_j\}$ je systém polynomů ortogonálních s vahou w na intervalu $[a, b]$. Pak platí:*

1. *Nechť $x_1 < x_2 < \ldots < x_n$ jsou kořeny polynomu p_n, $x_0 = a$, $x_{n+1} = b$. Pak každý interval $[x_k, x_{k+1}]$, $k = 0, 1, \ldots, n$ obsahuje právě jeden kořen polynomu p_{n+1}.*
2. *$p'_n(x)\, p_{n-1}(x) - p'_{n-1}(x)\, p_n(x) > 0 \ \forall x \in \mathbb{R}$, $n \geq 1$.*

Důkaz lze najít v [22].

Cvičení ke kapitole 8

1. Ukažte, že funkce $\phi_0 = 1/\sqrt{2\pi}$, $\phi_1(x) = (1/\sqrt{\pi}) \cos x$, $\ldots$, $\phi_n(x) = (1/\sqrt{\pi}) \cos nx$, $\phi_{n+1}(x) = (1/\sqrt{\pi}) \sin x$, $\ldots$, $\phi_{2n-1}(x) = (1/\sqrt{\pi}) \sin(n-1)x$ jsou ortogonální na $[-\pi, \pi]$ s vahou $w(x) \equiv 1$.

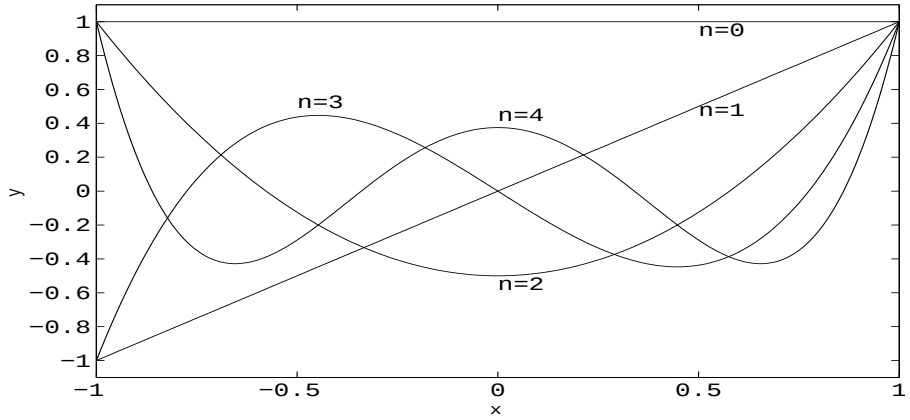
   (Návod: užijte trigonometrických identit pro $\cos(mx \pm nx)$, $\sin(mx \pm nx)$.)

2. Dokažte vztahy (8.3).

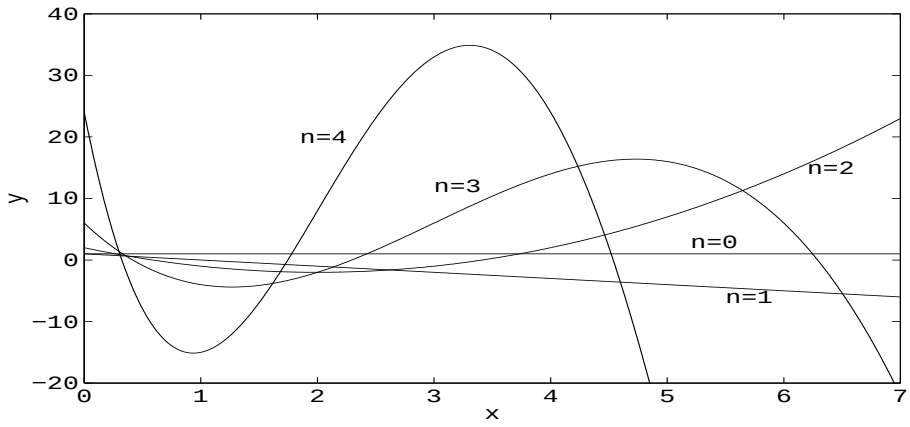
3. Dokažte, že pro každá přirozená čísla i, j platí

$$T_i(x)T_j(x) = \tfrac{1}{2}[T_{i+j}(x) + T_{|i-j|}(x)],$$

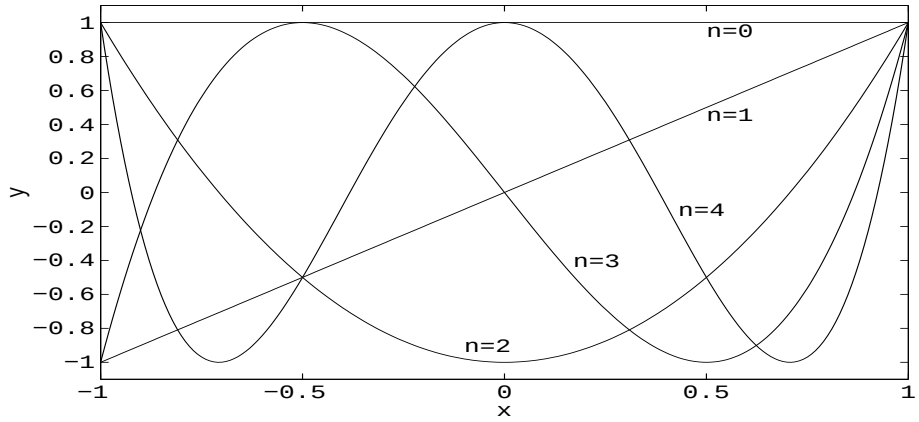
   kde T_i je Čebyševův polynom stupně i.



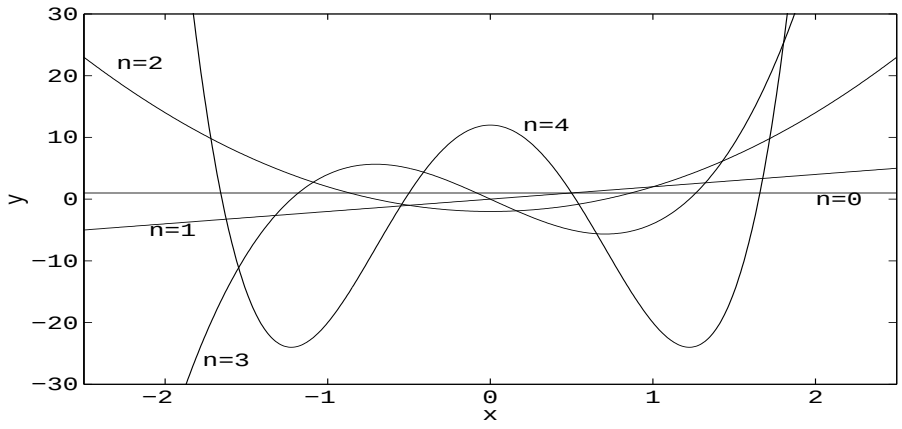
Obr. 8.1: Legendrovy polynomy P_n



Obr. 8.2: Laguerrovy polynomy L_n, $\alpha = 1$



Obr. 8.3: Čebyševovy polynomy T_n



Obr. 8.4: Hermitovy polynomy H_n

Kontrolní otázky ke kapitole 8

1. Nechť $\{\phi_0, \ldots, \phi_n\}$ je ortogonální systém funkcí na intervalu $[a, b]$ s vahou w. Je tento systém lineárně nezávislý?

2. Tvoří ortogonální polynomy $p_n, \ldots, p_0$ definované ve větě 8.1 Sturmovu posloupnost?

Kapitola 9

Numerické integrování

§ 9.1. Kvadraturní formule, stupeň přesnosti, chyba

Zabývejme se přibližným výpočtem integrálu

$$I(f) = \int_a^b f(x)\,\mathrm{d}x. \tag{9.1}$$

Z definice Riemannova integrálu a z jeho geometrického významu plyne, že je přirozené hledat aproximaci tohoto integrálu ve tvaru

$$I(f) \approx \sum_{i=0}^{n} A_i f(x_i), \tag{9.2}$$

kde body $x_i \in [a,b]$, $i = 0, 1, \ldots, n$, a reálná čísla A_i, $i = 0, 1, \ldots, n$, nezávisí na f.

Definice 9.1. Výraz

$$Q(f) = \sum_{i=0}^{n} A_i f(x_i) \tag{9.3}$$

budeme nazývat *kvadraturní formulí*, čísla A_i, $i = 0, 1, \ldots, n$, *koeficienty* kvadraturní formule a navzájem různé body x_i, $i = 0, 1, \ldots, n$, *uzly* kvadraturní formule.

Poznámka 1. Mějme integrál

$$I \ = \int_{-1}^{1} \frac{\mathrm{d}x}{(1+x^2)\sqrt{1-x^2}}.$$

Počítejme tento integrál pomocí formule (9.2). V případě, že body -1, 1 jsou uzly kvadraturní formule, nemůžeme tuto formuli použít, neboť integrand má singularity v těchto bodech. V tomto případě lze vhodně vyjádřit integrand pomocí vahové funkce (kap. 8) takto

$$I \ = \int_{-1}^{1} \frac{w(x)}{(1+x^2)}\,\mathrm{d}x, \qquad w(x) = \frac{1}{\sqrt{1-x^2}}.$$

Pro výpočet tohoto integrálu užijeme opět formule tvaru

$$\int_{-1}^{1} \frac{w(x)}{(1+x^2)}\,\mathrm{d}x \approx \sum_{i=0}^{n} A_i \frac{1}{(1+x_i^2)}.$$

Singularitu integrandu jsme tedy zahrnuli do vahové funkce, jejíž funkční hodnoty nevystupují explicitně v kvadraturní formuli. Myšlenku použití vahových funkcí v integrandu lze zobecnit a z tohoto důvodu se budeme obecně zabývat problémem aproximace integrálu

$$I(f) = \int_a^b w(x) f(x)\,\mathrm{d}x \tag{9.4}$$

formulí tvaru (9.3), tj.

$$\int_a^b w(x) f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} A_i f(x_i). \tag{9.5}$$

Vahovou funkci je rovněž vhodné zavést v případech, kdy počítáme celou řadu podobných integrálů a do funkce w můžeme zahrnout část společnou všem integrandům. Kvadraturní formule jsou většinou odvozeny integrací interpolačního polynomu a z tohoto důvodu je vhodné vybrat vahovou funkci tak, aby funkci f bylo možné dobře aproximovat interpolačním polynomem.

Ve formuli (9.5) vystupují uzly a koeficienty. Vzniká tedy otázka, jak vybrat tyto parametry. Jaké je kritérium přesnosti, tj. pro které funkce nastane ve vztahu (9.2) resp. (9.5) rovnost? Jaká je chyba této aproximace?

Rozdíl

$$R(f) = \int_a^b w(x) f(x)\,\mathrm{d}x - \sum_{i=0}^{n} A_i f(x_i) \tag{9.6}$$

budeme nazývat *chybou kvadraturní formule*.

Obecný tvar chyby zde nebudeme uvádět, ale u každé kvadraturní formule, kterou se budeme zabývat, uvedeme příslušný tvar chyby. Elegantní důkaz obecného tvaru chyby podal G. Peano (viz např [5], [8]).

Definice 9.2. Řekneme, že kvadraturní formule

$$Q(f) = \sum_{i=0}^{n} A_i f(x_i)$$

má *stupeň přesnosti* N, jestliže

$$R(x^j) = 0, \qquad j = 0, 1, \ldots, N, \qquad R(x^{N+1}) \neq 0. \tag{9.7}$$

Poznámka 2. V této definici se jedná o algebraický stupeň přesnosti, ne o chybu kvadraturní formule.

Příklad 9.1. Určete stupeň přesnosti kvadraturní formule $Q(f) = f(\alpha) + f(-\alpha)$, $0 < \alpha \leq 1$ pro výpočet integrálu $\int_{-1}^{1} f(x)\mathrm{d}x$.

Řešení: Pro funkci $f(x) = x^0$ je

$$R(x^0) = \int_{-1}^{1} 1\mathrm{d}x - (1-1) = 0.$$

Je-li $f(x) = x$, pak

$$R(x) = \int_{-1}^{1} x\mathrm{d}x - (\alpha - \alpha) = 0.$$

Dále

$$R(x^2) = \int_{-1}^{1} x^2\mathrm{d}x - (\alpha^2 + (-\alpha)^2) = \frac{2}{3} - 2\alpha^2.$$

Je-li $\alpha^2 \neq \frac{1}{3}$, pak má kvadraturní formule stupeň přesnosti roven jedné. Pro $\alpha = \frac{\sqrt{3}}{3}$ je

$$R(x^3) = \int_{-1}^{1} x^3\mathrm{d}x - ((\frac{\sqrt{3}}{3})^3 + (-\frac{\sqrt{3}}{3})^3) = 0,$$

$$R(x^4) = \int_{-1}^{1} x^4\mathrm{d}x - ((\frac{\sqrt{3}}{3})^4 + (-\frac{\sqrt{3}}{3})^4) \neq 0.$$

Pro $\alpha = \frac{\sqrt{3}}{3}$ je stupeň přesnosti kvadraturní formule roven třem.

Věta 9.1. *Kvadraturní formule užívající $n+1$ uzlů má stupeň přesnosti nejvýše $2n+1$.*

Důkaz. Předpokládejme, že kvadraturní formule

$$Q(f) = \sum_{i=0}^{n} A_i f(x_i)$$

pro výpočet integrálu (9.4) má stupeň přesnosti $2n+2$. Nechť x_i, $i = 0, 1, \ldots, n$, jsou uzly kvadraturní formule. Položme $\omega_{n+1}^2(x) = (x - x_0)^2 \ldots (x - x_n)^2$, ω_{n+1}^2 je polynom stupně $2n+2$. Počítejme nyní chybu kvadraturní formule pro výpočet integrálu $\int_a^b \omega_{n+1}^2(x) w(x)\,\mathrm{d}x$:

$$R(\omega_{n+1}^2) = \int_a^b w(x)\,\omega_{n+1}^2(x)\,\mathrm{d}x - \sum_{i=0}^{n} A_i \omega_{n+1}^2(x_i) = \int_a^b \omega_{n+1}^2(x) w(x)\,\mathrm{d}x,$$

neboť $\omega_{n+1}^2(x_i) = 0$, $i = 0, 1, \ldots, n$. Jelikož předpokládáme, že $N = 2n + 2$, plyne z tohoto vztahu:

$$R(\omega_{n+1}^2) = \int_a^b \omega_{n+1}^2(x) w(x)\,\mathrm{d}x = 0.$$

Toto je spor, neboť integrál z nezáporné funkce se nemůže rovnat nule. Stupeň přesnosti kvadraturní formule je tedy nejvýše $2n + 1$. □

Poznámka 3. Funkce uvedená v příkladu 9.1 má pro $\alpha = \sqrt{3}/3$ maximální stupeň přesnosti.

Věta 9.2. *Kvadraturní formule získaná integrací interpolačního polynomu určeného body $(x_i, f(x_i))$, $i = 0, \ldots, n$, má stupeň přesnosti alespoň n.*

Důkaz. Nechť $P_n \in \Pi_n$ je Lagrangeův tvar interpolačního polynomu funkce f v bodech x_i, $i = 0, 1, \ldots, n$:

$$P_n(x) = \sum_{i=0}^{n} l_i(x) f(x_i).$$

Pak

$$f(x) = P_n(x) + E(x) = \sum_{i=0}^{n} l_i(x) f(x_i) + \frac{\omega_{n+1}(x)}{(n+1)!} f^{(n+1)}(\xi), \qquad \xi = \xi(x).$$

Vynásobme tuto identitu vahovou funkcí w a integrujme v mezích od a do b:

$$\int_a^b w(x) f(x)\,\mathrm{d}x = \sum_{i=0}^{n} f(x_i) \int_a^b w(x) l_i(x)\,\mathrm{d}x + \\ + \frac{1}{(n+1)!} \int_a^b w(x)\, \omega_{n+1}(x) f^{(n+1)}(\xi)\,\mathrm{d}x.$$

Položme

$$A_i = \int_a^b w(x) l_i(x)\,\mathrm{d}x, \qquad i = 0, 1, \ldots, n.$$

Z předchozího vztahu plyne, že integrál

$$I(f) = \int_a^b w(x) f(x)\,\mathrm{d}x$$

je aproximován kvadraturní formulí

$$Q(f) = \sum_{i=0}^{n} A_i f(x_i)$$

s chybou

$$R(f) = \frac{1}{(n+1)!} \int_a^b w(x)\, \omega_{n+1}(x) f^{(n+1)}(\xi)\,\mathrm{d}x.$$

Ve vyjádření chyby vystupuje $(n+1)$ní derivace funkce f, tzn. že chyba kvadraturní formule bude rovna nule pro funkce $1, x, \ldots, x^n$ a stupeň přesnosti bude alespoň n. □

Příklad 9.2. Ukážeme, že formuli v příkladě 9.1 lze získat integrací interpolačního polynomu:
Nechť $P_1 \in \Pi_1$ je interpolační polynom pro funkci f v uzlech $-\alpha$, α:

$$P_1(x) = \frac{x-\alpha}{-2\alpha} f(-\alpha) + \frac{x+\alpha}{2\alpha} f(\alpha)\,.$$

Integrací dostaneme

$$\int_{-1}^{1} P_1(x)\mathrm{d}x = \frac{f(-\alpha)}{-2\alpha}\int_{-1}^{1}(x-\alpha)\mathrm{d}x + \frac{f(\alpha)}{2\alpha}\int_{-1}^{1}(x+\alpha)\mathrm{d}x = f(-\alpha)+f(\alpha)\,.$$

Poznámka 4. Jsou-li dány uzly kvadraturní formule $x_0, \ldots, x_n$, $x_i \neq x_k$ pro $i \neq k$, můžeme vždy najít koeficienty $A_0, \ldots, A_n$ tak, aby stupeň přesnosti byl roven n, neboť systém rovnic

$$\sum_{i=0}^{n} A_i x_i^k = \int_a^b x^k w(x)\mathrm{d}x, \quad k = 0, \ldots, n$$

má právě jedno řešení. Pro dané uzly existuje tedy právě jedna posloupnost koeficientů $A_0, \ldots, A_n$. Na druhé straně můžeme tyto koeficienty získat integrací interpolačního polynomu a chyba kvadraturní formule je v takovém případě dána vztahem

$$R(f) = \frac{1}{(n+1)!}\int_a^b w(x)\omega_{n+1}(x) f^{(n+1)}(\xi)\mathrm{d}x\,.$$

V závislosti na vlastnostech uzlů a vahové funkce lze tento integrál dále upravit (např. metodou per partes) tak, že ve vyjádření chyby bude vystupovat $f^{(N+1)}(\eta)$, kde N je stupeň přesnosti dané formule.

Věta 9.3. *Nechť vahová funkce w je sudá vzhledem ke středu s intervalu $[a,b]$ a nechť uzly x_i, $i = 0, 1, \ldots, n$, jsou symetricky rozloženy vzhledem ke středu s. Pak koeficienty kvadraturní formule (získané integrací interpolačního polynomu v uzlech x_i, $i = 0, \ldots, n$) odpovídající symetrickým uzlům jsou stejné, tj.*

$$A_i = A_{n-i}, \qquad i = 0, 1, \ldots, n. \tag{9.8}$$

Důkaz lze najít např. v [1].

Uvažujme takovou „symetrickou“ kvadraturní formuli. Tato formule bude přesná pro libovolnou funkci f lichou vzhledem ke středu s intervalu $[a, b]$. Pro takovou funkci je totiž

$$\int_a^b w(x)f(x)\,\mathrm{d}x = 0$$

a na druhé straně

$$Q(f) = \sum_{i=0}^{n} A_i f(x_i) = 0,$$

neboť $f(x_i) = -f(x_{n-i})$, $A_i = A_{n-i}$, $i = 0, 1, \ldots, n$.

Nechť nyní $(n+1)$, tj. počet uzlů této symetrické kvadraturní formule, je číslo liché. Tato formule bude zřejmě přesná i pro polynom

$$P_{n+1}(x) = \left(x - \frac{a+b}{2}\right)^{n+1},$$

neboť P_{n+1} je lichá funkce vzhledem ke středu $s = (a+b)/2$. Podle předchozí věty víme, že kvadraturní formule odvozená integrací interpolačního polynomu má stupeň přesnosti alespň n. Ukážeme nyní, že tato symetrická kvadraturní formule má pro $(n+1)$ liché stupeň přesnosti alespoň $n+1$. Je totiž

$$\int_a^b P_{n+1}(x)w(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i P_{n+1}(x_i),$$

kde $P_{n+1}(x) = (x - \frac{1}{2}(a+b))^{n+1}$.

Dále

$$\int_a^b P_{n+1}(x)w(x)\,\mathrm{d}x = \int_a^b x^{n+1}w(x)\,\mathrm{d}x - \int_a^b \left(\sum_{j=0}^{n} d_j x^j\right) w(x)\,\mathrm{d}x =$$
$$= \sum_{i=0}^{n} A_i x_i^{n+1} - \sum_{i=0}^{n} A_i \left(\sum_{j=0}^{n} d_j x_i^j\right).$$

Stupeň přesnosti kvadraturní formule je alespoň n, a proto

$$\int_a^b \left(\sum_{j=0}^{n} d_j x^j\right) w(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i \left(\sum_{j=0}^{n} d_j x_i^j\right).$$

A odtud plyne

$$\int_a^b x^{n+1}w(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i x_i^{n+1},$$

a to znamená, že stupeň přesnosti kvadraturní formule je alespoň $n+1$.

Zabývejme se nyní otázkou výběru uzlů a koeficientů kvadraturní formule. V praxi jsou často uzly dány vnějšími okolnostmi nebo se užívá ekvidistantních uzlů. Ale na volbu uzlů a koeficientů se také můžeme dívat z hlediska stupně přesnosti. Formulujme nyní nejčastější požadavky na uzly a koeficienty kvadraturní formule:

a) nejsou předem dána žádná omezení ani na uzly ani na koeficienty kvadraturní formule,

b) jsou předepsány všechny uzly, obvykle ekvidistantní,

c) jsou předepsány pouze některé uzly, např. koncové body intervalu nebo body, v nichž je chování funkce význačné,

d) požaduje se rovnost všech koeficientů.

§ 9.2. Gaussovy kvadraturní formule

Zabývejme se nejdříve případem, kdy nejsou dána žádná omezení ani na uzly ani na koeficienty kvadraturní formule. Je zde tedy problém: můžeme vybrat uzly a koeficienty tak, aby bylo dosaženo maximálního stupně přesnosti?

Na tuto otázku dává odpověď následující věta:

Věta 9.4. *Nechť kvadraturní formule*

$$Q(f) = \sum_{i=0}^{n} A_i f(x_i)$$

pro výpočet integrálu (9.4) *má stupeň přesnosti alespoň* n. *Nechť* $\{p_n\}$, $p_n \in \bar{\Pi}_n$, $n = 0, 1, \ldots$ *tvoří ortogonální systém na intervalu* $[a, b]$ *vzhledem k vahové funkci* w. *Pak tato formule má stupeň přesnosti* $2n+1$ *právě tehdy, když uzly této kvadraturní formule jsou kořeny polynomu* $p_{n+1} \in \bar{\Pi}_{n+1}$.

Důkaz. Nechť kvadraturní formule $Q(f)$ má stupeň přesnosti $2n + 1$. Definujme polynom $\omega_{n+1} \in \bar{\Pi}_{n+1}$ vztahem

$$\omega_{n+1}(x) = \prod_{i=0}^{n} (x - x_i),$$

kde x_i, $i = 0, 1, \ldots, n$, jsou uzly dané kvadraturní formule. Podle předpokladu je kvadraturní formule přesná i pro polynom $u_n \omega_{n+1}$, kde $u_n \in \Pi_n$ je libovolný polynom.

Je tedy

$$\int_a^b \omega_{n+1}(x) u_n(x) w(x)\, \mathrm{d}x = \sum_{i=0}^{n} A_i\, \omega_{n+1}(x_i) u_n(x_i).$$

Pravá strana této rovnosti je rovna nule, neboť $\omega_{n+1}(x_i) = 0$, $i = 0, 1, \ldots, n$, a tedy pro libovolný polynom $u_n \in \Pi_n$ platí

$$\int_a^b w(x)\, \omega_{n+1}(x) u_n(x)\, \mathrm{d}x = 0.$$

Polynom ω_{n+1} je ortogonální s vahou w ke všem polynomům ze třídy Π_n, tzn. že polynom ω_{n+1} je totožný s polynomem p_{n+1}, který náleží systému polynomů $\{p_n\}$ ortogonálních s vahou w na intervalu $[a, b]$, $p_n \in \bar{\Pi}_n$, (věta 8.1 a její důsledek).

Uvažujme nyní systém $\{p_n\}$ polynomů ortogonálních s vahou w na intervalu $[a, b]$. Sestrojme kvadraturní formuli, jejíž uzly jsou kořeny ortogonálního polynomu p_{n+1}. Z věty 8.2 víme, že všechny kořeny tohoto polynomu jsou reálné, navzájem různé a všechny leží v (a, b).

Uvažujme formuli mající stupeň přesnosti alespoň n. Takovou formuli lze snadno sestrojit integrací interpolačního polynomu (viz věta 9.2). Nechť $Q(f)$ je taková formule a ukážeme, že má stupeň přesnosti $2n + 1$.

Nechť $P_{2n+1} \in \Pi_{2n+1}$ je libovolný polynom stupně $2n + 1$. Zřejmě můžeme tento polynom zapsat ve tvaru

$$P_{2n+1}(x) = p_{n+1}(x) u_n(x) + r_n(x),$$

kde u_n je podíl při dělení P_{2n+1}/p_{n+1} a r_n je zbytek při tomto dělení. Je zřejmě $r_n, u_n \in \Pi_n$.

Aplikujme nyní na P_{2n+1} sestrojenou kvadraturní formuli a vypočtěme chybu aproximace integrálu

$$\begin{aligned} R(P_{2n+1}) &= \int_a^b w(x) P_{2n+1}(x)\, \mathrm{d}x - \sum_{i=0}^{n} A_i P_{2n+1}(x_i) = \\ &= \left\{ \int_a^b w(x) u_n(x) p_{n+1}(x)\, \mathrm{d}x - \sum_{i=0}^{n} A_i u_n(x_i) p_{n+1}(x_i) \right\} + \\ &\quad + \left\{ \int_a^b w(x) r_n(x)\, \mathrm{d}x - \sum_{i=0}^{n} A_i r_n(x_i) \right\}. \end{aligned}$$

Výraz v první závorce je roven nule, neboť polynom p_{n+1} je ortogonální s vahou w k polynomu u_n a navíc $p_{n+1}(x_i) = 0$, $i = 0, \ldots, n$, neboť uzly kvadraturní formule jsou kořeny polynomu p_{n+1}. Výraz v druhé závorce je rovněž roven nule, neboť stupeň přesnosti kvadraturní formule je alespoň n. Odtud plyne $R(P_{2n+1}) = 0$ pro libovolný polynom z Π_{2n+1} a tedy stupeň přesnosti kvadraturní formule je $2n + 1$. □

Definice 9.3. Kvadraturní formule, jejichž uzly a koeficienty jsou vybrány tak, aby bylo dosaženo maximálního stupně přesnosti, se nazývají *Gaussovy kvadraturní formule*.

Uvedená věta uvádí podmínky pro uzly kvadraturní formule. Vlastnosti koeficientů této kvadraturní formule jsou obsaženy implicitně v předpokladu, že stupeň

přesnosti je alespoň n. Formule se stupněm přesnosti n můžeme snadno sestrojit integrací interpolačního polynomu (věta 9.2). Koeficienty můžeme tedy spočítat takto (viz poznámka 4):

$$A_i = \int_a^b w(x) l_i(x)\,\mathrm{d}x, \qquad i = 0, 1, \ldots, n, \tag{9.9}$$

kde x_i, $i = 0, 1, \ldots, n$, jsou kořeny ortogonálního polynomu, $p_{n+1}(x) \equiv \omega_{n+1}(x)$,

$$l_i(x) = \frac{p_{n+1}(x)}{p'_{n+1}(x_i)(x - x_i)}, \qquad i = 0, 1, \ldots, n,$$

jsou příslušné fundamentální polynomy v Lagrangeově interpolačním polynomu.

Jiný způsob výpočtu koeficientů vychází přímo z definice stupně přesnosti. Požadujeme-li totiž, aby kvadraturní formule měla stupeň přesnosti n, pak musí být splněny podmínky:

$$\int_a^b w(x) x^k\,\mathrm{d}x = \sum_{i=0}^{n} A_i x_i^k, \qquad k = 0, 1, \ldots, n. \tag{9.10}$$

Víme, že všechny uzly x_i, $i = 0, 1, \ldots, n$, jsou reálné, různé a všechny leží v (a, b). Odtud plyne, že determinant soustavy, tzv. Vandermondův, je různý od nuly a soustava má jediné řešení, a tedy Gaussova kvadraturní formule je určena jednoznačně.

Věta 9.5. *Pro koeficienty Gaussovy kvadraturní formule platí*

$$a)\ A_i > 0, \quad i = 0, 1, \ldots, n, \qquad b)\ \sum_{i=0}^{n} A_i = \int_a^b w(x)\,\mathrm{d}x.$$

Důkaz.

a) Gaussovy kvadraturní formule mají stupeň přesnosti $2n + 1$ a jsou přesné i pro polynomy l_j^2, $j = 0, 1, \ldots, n$, což jsou polynomy stupně $2n$:

$$l_j^2(x) = \left(\frac{p_{n+1}(x)}{p'_{n+1}(x_j)(x - x_j)}\right)^2.$$

Pro výpočet integrálu $I(l_j^2) = \int_a^b w(x) l_j^2(x)\,\mathrm{d}x$ tedy platí

$$\int_a^b w(x) l_j^2(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i l_j^2(x_i).$$

Jelikož $l_j(x_i) = \delta_{ij}$, $i, j = 0, 1, \ldots, n$, plyne odtud

$$A_j = \int_a^b w(x) l_j^2(x)\,\mathrm{d}x > 0, \qquad j = 0, 1, \ldots, n. \tag{9.11}$$

Navíc porovnáním vztahů (9.9) a (9.11) dostáváme zajímavou rovnost

$$\int_a^b w(x)l_j^2 \,\mathrm{d}x = \int_a^b w(x)l_j(x) \,\mathrm{d}x, \qquad j = 0, 1, \ldots, n. \tag{9.12}$$

b) Aplikací Gaussovy kvadraturní formule na funkci $f(x) \equiv 1$, pro kterou je samozřejmě kvadraturní formule přesná, ihned dostaneme

$$\int_a^b w(x) \,\mathrm{d}x = \sum_{i=0}^{n} A_i.$$

□

Věta 9.6. *Nechť $f \in C^{(2n+2)}[a, b]$. Chybu Gaussovy kvadraturní formule lze vyjádřit ve tvaru*

$$R(f) = \frac{f^{(2n+2)}(\eta)}{(2n+2)!} \int_a^b w(x)p_{n+1}^2(x) \,\mathrm{d}x, \qquad \eta \in (a, b), \tag{9.13}$$

kde $\omega_{n+1}(x) = p_{n+1}(x)\ \forall\ x$.

Důkaz. Funkci f vyjádříme pomocí Hermitova interpolačního polynomu $P_{2n+1} \in \Pi_{2n+1}$ splňujícího podmínky

$$P_{2n+1}(x_i) = f(x_i), \qquad i = 0, 1, \ldots, n,$$
$$P'_{2n+1}(x_i) = f'(x_i), \qquad i = 0, 1, \ldots, n,$$

kde x_i, $i = 0, 1, \ldots, n$, jsou uzly Gaussovy kvadraturní formule. Podle kapitoly 8 víme, že funkci f lze pomocí tohoto polynomu vyjádřit ve tvaru

$$f(x) = P_{2n+1}(x) + E(x),$$

kde

$$P_{2n+1}(x) = \sum_{i=0}^{n} h_i(x)f(x_i) + \sum_{i=0}^{n} \bar{h}_i(x)f'(x_i),$$
$$E(x) = \frac{p_{n+1}^2(x)}{(2n+2)!} f^{(2n+2)}(\xi), \qquad \xi = \xi(x).$$

Připomínáme, že v tomto případě $\omega_{n+1}(x) = p_{n+1}(x)$, p_{n+1} je ortogonální polynom (s vahou w na $[a, b]$).

Užijeme nyní tohoto vyjádření při výpočtu chyby kvadraturní formule:

$$R(f) = \int_a^b w(x)f(x) \,\mathrm{d}x - \sum_{i=0}^{n} A_i f(x_i) =$$
$$= \left\{ \int_a^b P_{2n+1}(x)w(x) \,\mathrm{d}x - \sum_{i=0}^{n} A_i P_{2n+1}(x_i) \right\} +$$
$$+ \left\{ \int_a^b E(x)w(x) \,\mathrm{d}x - \sum_{i=0}^{n} A_i E(x_i) \right\}.$$

Výraz v první závorce je roven nule, neboť formule má stupeň přesnosti $2n+1$; dále $E(x_i)=0$, $i=0,1,\dots,n$, neboť $p_{n+1}(x_i)=0$, $i=0,1,\dots,n$. Odtud plyne

$$R(f)=\int_a^b \frac{w(x)p_{n+1}^2(x)}{(2n+2)!}f^{(2n+2)}(\xi)\,\mathrm{d}x.$$

Funkce $f\in C^{(2n+2)}[a,b]$ a funkce wp_{n+1}^2 je nezáporná a integrovatelná v $[a,b]$. Lze tedy užít druhé věty o střední hodnotě integrálu[1]. Výsledkem je následující tvar chyby Gaussovy kvadraturní formule:

$$R(f)=\frac{f^{(2n+2)}(\eta)}{(2n+2)!}\int_a^b w(x)p_{n+1}^2(x)\,\mathrm{d}x,\qquad \eta\in(a,b).$$

□

Popišme nyní podrobněji konstrukci Gaussových formulí pro případ $[a,b]=[-1,1]$ a $w(x)=1\ \forall x\in[-1,1]$. To znamená, že hledáme Gaussovu formuli pro výpočet integrálu

$$\int_{-1}^1 f(x)\,\mathrm{d}x.$$

Polynomy, které jsou ortogonální na intervalu $[-1,1]$ s vahou $w(x)\equiv 1$, se nazývají Legendrovy polynomy (viz kap. 8) a označujeme je P_n. Z rekurentního vztahu

$$P_{n+1}(x)=\frac{2n+1}{n+1}xP_n(x)-\frac{n}{n+1}P_{n-1}(x),\qquad n=1,2,\dots$$

$P_0(x)=0$, $P_1(x)=x$, plyne, že je-li stupeň polynomu P_n číslo sudé, obsahuje P_n pouze sudé mocniny x, je-li stupeň polynomu číslo liché, obsahuje polynom pouze liché mocniny x. Odtud dále plyne, kořeny polynomu P_n jsou symetricky rozloženy vzhledem k bodu 0. Uzly x_i, $i=0,1,\dots,n$, Gaussovy kvadraturní formule jsou kořeny Legendreova polynomu P_{n+1} a koeficienty lze spočítat takto ([19]):

$$A_i=\frac{2(1-x_i)^2}{(n+1)^2(P_n(x_i))^2},\qquad i=0,1,\dots,n. \tag{9.14}$$

Výsledná formule je tvaru

$$\int_{-1}^1 f(x)\,\mathrm{d}x=\sum_{i=0}^n A_i f(x_i)+\frac{f^{(2n+2)}(\alpha)}{(2n+2)!}\int_{-1}^1\prod_{i=0}^n(x-x_i)^2\,\mathrm{d}x,\quad \alpha\in(-1,1). \tag{9.15}$$

Tato formule se nazývá *Gaussova-Legendreova* kvadraturní formule.

[1] Druhá věta o střední hodnotě integrálu – někdy se také nazývá zobecněná věta o střední hodnotě integrálu: Nechť $\varphi\in C[a,b]$, ψ integrovatelná v $[a,b]$ a nemění znaménko v $[a,b]$. Pak existuje bod $\tau\in(a,b)$, že

$$\int_a^b \varphi(x)\psi(x)\,\mathrm{d}x=\varphi(\tau)\int_a^b\psi(x)\,\mathrm{d}x.$$

Legendrovy polynomy uvedené v předchozí kapitole splňují vztah

$$\int_{-1}^{1} P_n^2(x)\,\mathrm{d}x = \frac{2}{2n+1}.$$

Tyto polynomy nemají koeficient 1 u nejvyšší mocniny. Položme nyní

$$p_n(x) = \frac{1}{a_n} P_n(x),$$

kde a_n je koeficient u x^n v polynomu P_n. Pak

$$\frac{2}{2n+1} = \int_{-1}^{1} P_n^2(x)\,\mathrm{d}x = a_n^2 \int_{-1}^{1} p_n^2(x)\,\mathrm{d}x.$$

Odtud

$$\int_{-1}^{1} p_n^2(x)\,\mathrm{d}x = \frac{2}{(2n+1)a_n^2}.$$

Lze ukázat [19], [21], že

$$a_n = \frac{(2n)!}{2^n (n!)^2}.$$

A z toho plyne, že chyba Gaussovy-Legendrovy formule může být vyjádřena ve tvaru

$$R(f) = \frac{f^{(2n+2)}(\alpha)}{(2n+2)!} \frac{2}{2(n+1)+1} \left(\frac{2^{n+1}((n+1)!)^2}{(2(n+1))!} \right)^2.$$

Podobným postupem lze vyjádřit chybu i pro další kvadraturní formule.

Výpočet integrálu na libovolném intervalu $[a,b]$ lze vhodnou substitucí převést na výpočet na intervalu $[-1,1]$ a poté použít formule (9.15).

Příklad 9.3. Odvoďte Gaussovu-Legendrovu formuli ve tvaru

$$\int_{-1}^{1} f(x)\mathrm{d}x \approx A_0 f(x_0) + A_1 f(x_1)\,.$$

Řešení. Legendrův polynom druhého stupně je tvaru

$$P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}\,.$$

Jeho kořeny jsou $x_0 = -\sqrt{3}/3$, $x_1 = \sqrt{3}/3$, a to jsou dva uzly kvadraturní formule.

Nejprve spočítáme koeficienty kvadraturní formule integrací interpolačního polynomu. Užijeme vzorce (9.9), kde $w(x) \equiv 1$ a

$$l_0(x) = \frac{x - \frac{\sqrt{3}}{3}}{-\frac{2\sqrt{3}}{3}}, \quad l_1(x) = \frac{x + \frac{\sqrt{3}}{3}}{\frac{2\sqrt{3}}{3}}\,.$$

Koeficienty

$$A_0 = \int_{-1}^{1} l_0(x)\mathrm{d}x = 1, \quad A_1 = \int_{-1}^{1} l_1(x)\mathrm{d}x = 1\,.$$

Výsledná formule je tedy tvaru

$$\int_{-1}^{1} f(x)\mathrm{d}x = f(-\frac{\sqrt{3}}{3}) + f(\frac{\sqrt{3}}{3}) + R(f)\,. \tag{9.16}$$

Samozřejmě, že koeficienty A_0, A_1 můžeme určit také řešením systému (9.10):

$$\begin{aligned} A_0 + A_1 &= \int_{-1}^{1} 1\mathrm{d}x = 2 \\ -A_0\frac{\sqrt{3}}{3} + A_1\frac{\sqrt{3}}{3} &= \int_{-1}^{1} x\mathrm{d}x = 0 \end{aligned}$$

Odtud pak ihned plyne, že $A_0 = A_1 = 1$.

Příklad 9.4. Ukážeme ještě jeden způsob konstrukce Gaussových formulí. Uvažujme formuli $Q(f) = A_0 f(x_0) + A_1 f(x_1)$ pro výpočet integrálu $\int_{-1}^{1} f(x)\mathrm{d}x$. Víme, že taková kvadraturní formule může mít nejvýše stupeň přesnosti $2n+1$, což je v našem případě 3 (viz věta 9.1). Podmínky pro dosažení tohoto stupně přesnosti jsou dány rovnicemi $R(x^k) = 0$, $k = 0, 1, 2, 3$, tj.

$$\begin{aligned} A_0 + A_1 &= \int_{-1}^{1} 1\mathrm{d}x = 2 \\ A_0x_0 + A_1x_1 &= \int_{-1}^{1} x\mathrm{d}x = 0 \\ A_0x_0^2 + A_1x_1^2 &= \int_{-1}^{1} x^2\mathrm{d}x = \frac{2}{3} \\ A_0x_0^3 + A_1x_1^3 &= \int_{-1}^{1} x^3\mathrm{d}x = 0\,. \end{aligned}$$

Uzly x_0, x_1 jsou kořeny polynomu $\omega_2(x) = (x - x_0)(x - x_1) = x^2 + a_1x + a_2$. Je třeba určit koeficienty A_1, A_2. Postup je následující:

Vynásobíme první z výše uvedených rovnic a_2, druhou a_1 třetí jedničkou a sečteme je. Pak druhou rovnici vynásobíme a_2, třetí a_1 čtvrtou jedničkou a opět sečteme:

$$\begin{aligned} a_2(A_0+A_1)+a_1(A_0x_0+A_1x_1)+A_0x_0^2+A_1x_1^2 &= 2a_2+\frac{2}{3} \\ a_2(A_0x_0+A_1x_1)+a_1(A_0x_0^2+A_1x_1^2)+A_0x_0^3+A_1x_1^3 &= \frac{2}{3}a_1\,. \end{aligned}$$

Víme, že $x_i^2+a_1x_i+a_2=0$, $i=0,\,1$, a proto jsou levé strany těchto rovnic rovny nule:

$$\begin{aligned} 0 &= 2a_2+\frac{2}{3} \Rightarrow a_2=-\frac{1}{3} \\ 0 &= a_1\,. \end{aligned}$$

Pak polynom $\omega_2(x)=x^2-\frac{1}{3}$, odtud $x_0=-\frac{\sqrt{3}}{3}$, $x_1=\frac{\sqrt{3}}{3}$. Koeficienty A_0, A_1 získáme z prvních dvou rovnic: $A_0=A_1=1$.

Poznámka 5. Víme, že chyba Gaussovy kvadraturní formule může být vyjádřena ve tvaru

$$R(f)=\frac{f^{(2n+2)}(\eta)}{(2n+2)!}\int_a^b w(x)p_{n+1}^2(x)\,\mathrm{d}x.$$

Veličina $\int_a^b w(x)p_{n+1}^2(x)\,\mathrm{d}x$ nezávisí na volbě funkce f. Je tedy pro všechny funkce stejná. Vypočtěme nyní chybu kvadraturní formule pro funkci $f(x)=x^{2n+2}$. Tato chyba je tvaru

$$R(x^{2n+2})=\int_a^b w(x)x^{2n+2}\,\mathrm{d}x-\sum_{i=0}^{n}A_ix_i^{2n+2}.$$

Z výše uvedeného vztahu plyne, že pro $f(x)=x^{2n+2}$ je

$$R(x^{2n+2})=\frac{(2n+2)!}{(2n+2)!}\int_a^b w(x)p_{n+1}^2(x)\,\mathrm{d}x$$

a odtud

$$\int_a^b w(x)p_{n+1}^2(x)\,\mathrm{d}x=R(x^{2n+2}).$$

Výraz pro chybu Gaussovy kvadraturní formule můžeme zapsat takto:

$$R(f)=\frac{f^{(2n+2)}(\eta)}{(2n+2)!}R(x^{2n+2}).$$

Tedy např. pro Gaussovu-Legendrovu formuli pro $n=1$ je

$$R(f)=\frac{f^{(4)}(\alpha)}{4!}R(x^4),\ \alpha\in(-1,1)$$

kde

$$R(x^4) = \int_{-1}^{1} x^4 \,\mathrm{d}x - \left(\left(-\frac{\sqrt{3}}{3} \right)^4 + \left(\frac{\sqrt{3}}{3} \right)^4 \right) = \frac{8}{45}$$

a odtud

$$R(f) = \frac{1}{135} f^{(4)}(\alpha).$$

Poznámka 6. Nejjednodušší Gaussova-Legendreova formule je tvaru:

$$\int_{-1}^{1} f(x)\mathrm{d}x = 2f(0) + \frac{f''(\beta)}{3}\,. \tag{9.17}$$

Zde totiž $n = 0$, Legendrův polynom $P_1(x) = x$ má kořen $x_0 = 0$ a pak

$$A_0 = \int_{-1}^{1} l_0(x)\mathrm{d}x = \int_{-1}^{1} 1\mathrm{d}x = 2\,.$$

Výpočet chyby $R(f)$ v tomto případě:
Je $R(x^2) = 2/3$ a odtud

$$R(f) = \frac{1}{2}\frac{2}{3} f''(\beta) = \frac{1}{3} f''(\beta),\ \beta \in (-1, 1).$$

Jak už jsme se zmínili dříve, obecný tvar chyby kvadraturních formulí dokázal G. Peano. Ukážeme nyní na příkladě Gaussovy-Legendrovy formule myšlenku tohoto důkazu.

Mějme formuli

$$\int_{-1}^{1} f(x)\mathrm{d}x = f\left(-\frac{\sqrt{3}}{3} \right) + f\left(\frac{\sqrt{3}}{3} \right) + R(f)\,.$$

Jedná se o Gaussovu-Legendrovu formuli pro $n = 1$. Víme, že stupeň přesnosti této formule je $N = 3$. Předpokládejme, že $f \in C^4[-1, 1]$. Uvažujme Taylorův rozvoj funkce f v okolí bodu $x = -1$ ve tvaru

$$f(x) = f(-1) + f'(-1)(x+1) + \frac{f''(-1)}{2}(x+1)^2 + \frac{f'''(-1)}{x!}(x+1)^3 + r_3(x)\,,$$

kde zbytek $r_3(x)$ je zapsán v integrálním tvaru

$$r_3(x) = \frac{1}{3!} \int_{-1}^{x} f^{(4)}(t)(x-t)^3 \mathrm{d}t\,.$$

Funkci f můžeme nyní zapsat takto

$$f(x) = P_3(x) + r_3(x),\ P_3 \in \Pi_3\,.$$

Aplikujeme nyní na takto vyjádřenou funkci danou kvadraturní formuli a počítejme chybu:

$$R(f) = \int_{-1}^{1} f(x)\mathrm{d}x - \left(f\left(-\frac{\sqrt{3}}{3}\right) + f\left(\frac{\sqrt{3}}{3}\right)\right) =$$

$$= \left\{ \int_{-1}^{1} P_3(x)\mathrm{d}x - \left(P_3\left(-\frac{\sqrt{3}}{3}\right) + P_3\left(\frac{\sqrt{3}}{3}\right)\right)\right\} +$$

$$+ \left\{ \int_{-1}^{1} r_3(x)\mathrm{d}x - \left(r_3\left(-\frac{\sqrt{3}}{3}\right) + r_3\left(\frac{\sqrt{3}}{3}\right)\right)\right\}\,.$$

Výraz v první závorce je roven nule, neboť stupeň přesnosti kvadraturní formule je $N = 3$, a tedy

$$R(f) = \frac{1}{3!}\int_{-1}^{1}\left(\int_{-1}^{x} f^{(4)}(t)(x-t)^3\mathrm{d}t\right)\mathrm{d}x -$$

$$-\frac{1}{3!}\left(\int_{-1}^{-\frac{\sqrt{3}}{3}} f^{(4)}(t)\left(-\frac{\sqrt{3}}{3}-t\right))^3\mathrm{d}t + \int_{-1}^{\frac{\sqrt{3}}{3}} f^{(4)}(t)\left(\frac{\sqrt{3}}{3}-t\right))^3\mathrm{d}t\right)\,.$$

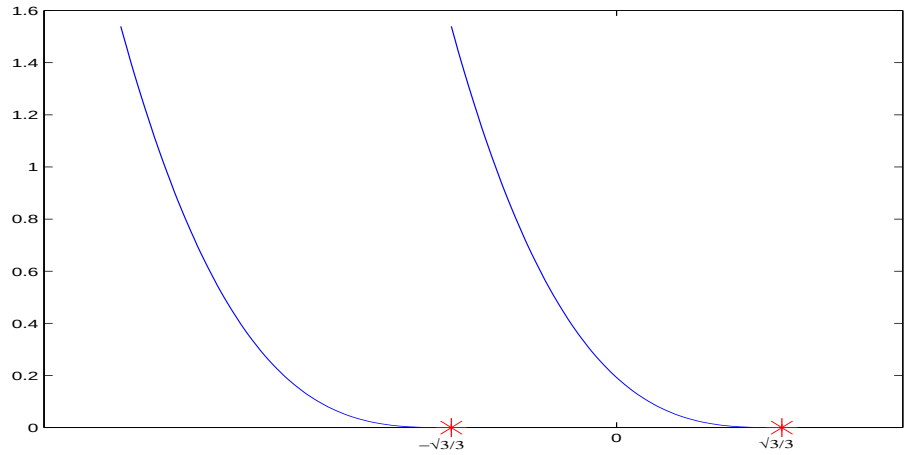
Definujme nyní funkci $(x-t)_+^3$ jako funkci proměnné t takto:

$$(x-t)_+^3 = \begin{cases} (x-t)^3 & x \geq t \\ 0 & x < t\,. \end{cases}$$

Na obrázku 9.1 vidíme průběh funkcí $(-\frac{\sqrt{3}}{3}-t)_+^3$ a $(\frac{\sqrt{3}}{3}-t)_+^3$. Tyto funkce se někdy nazývají „useknuté“ mocniny a zavedli jsme je proto, aby integrační meze v předchozím vyjádření byly konstanty, což pak umožní další výpočet. Máme tedy integrál

$$R(f) = \frac{1}{3!}\int_{-1}^{1}\left(\int_{-1}^{1} f^{(4)}(t)(x-t)_+^3\mathrm{d}t\right)\mathrm{d}x -$$

$$-\frac{1}{3!}\left(\int_{-1}^{1} f^{(4)}(t)\left(-\frac{\sqrt{3}}{3}-t\right)_+^3\mathrm{d}t + \int_{-1}^{1} f^{(4)}(t)\left(\frac{\sqrt{3}}{3}-t\right)_+^3\mathrm{d}t\right)\,.$$



Obr. 9.1: Funkce $(-\frac{\sqrt{3}}{3}-t)_+^3$ a $(\frac{\sqrt{3}}{3}-t)_+^3$.

Integrační meze jsou konstanty a můžeme zaměnit pořadí integrace v prvním integrálu.

$$R(f)=\frac{1}{3!}\int_{-1}^{1} f^{(4)}(t)\left(\int_{-1}^{1}(x-t)_+^3\mathrm{d}x\right)\mathrm{d}t-$$

$$-\frac{1}{3!}\left(\int_{-1}^{1} f^{(4)}(t)\left(-\frac{\sqrt{3}}{3}-t\right)_+^3\mathrm{d}t+\int_{-1}^{1} f^{(4)}(t)\left(\frac{\sqrt{3}}{3}-t\right)_+^3\mathrm{d}t\right)=$$

$$=\frac{1}{3!}\int_{-1}^{1} f^{(4)}(t)\left\{\int_{-1}^{1}(x-t)_+^3\mathrm{d}x-\left(-\frac{\sqrt{3}}{3}-t\right)_+^3-\left(\frac{\sqrt{3}}{3}-t\right)_+^3\right\}\mathrm{d}t\,.$$

Výraz

$$R_x((x-t)_+^3)=\int_{-1}^{1}(x-t)_+^3\mathrm{d}x-\left(-\frac{\sqrt{3}}{3}-t\right)_+^3-\left(\frac{\sqrt{3}}{3}-t\right)_+^3$$

je chyba dané kvadraturní formule pro funkci $\varphi(x)=(x-t)_+^4$.

Položme nyní $K(t)=\frac{1}{3!}R_x((x-t)_+^3)$ a chybu kvadraturní formule můžeme zapsat ve tvaru

$$R(f)=\int_{-1}^{1} f^{(4)}(t)K(t)\mathrm{d}t\,,$$

kde K je *Peanovo jádro*.

Podívejme se nyní na vlastnosti funkce K:

$$K(t) = \frac{1}{3!}\left(\int_{-1}^{1}(x-t)_+^3\mathrm{d}x - \left(-\frac{\sqrt{3}}{3}-t\right)_+^3 - \left(\frac{\sqrt{3}}{3}-t\right)_+^3\right).$$

Z vlastností funkce $(x-t)_+^3$ plyne, že

$$\int_{-1}^{1}(x-t)_+^3\mathrm{d}x = \int_{t}^{1}(x-t)^3\mathrm{d}x = \frac{(1-t)^4}{4}.$$

Pro funkci K platí:

$$K(t) = \frac{1}{3!}\left(\frac{(1-t)^4}{4} - \left(-\frac{\sqrt{3}}{3}-t\right)_+^3 - \left(\frac{\sqrt{3}}{3}-t\right)_+^3\right).$$

Vyšetřeme nyní chování této funkce postupně na intervalech $[-1,-\sqrt{3}/3]$, $[-\sqrt{3}/3,\sqrt{3}/3]$, $[\sqrt{3}/3,1]$.

(1) $t\in[-1,-\sqrt{3}/3] \Rightarrow K(t) = \frac{1}{3!}\left(\frac{(1-t)^4}{4} + \left(+\frac{\sqrt{3}}{3}+t\right)^3 - \left(\frac{\sqrt{3}}{3}-t\right)^3\right) = \frac{(1+t)^4}{24}$

(2) $t\in[-\sqrt{3}/3,\sqrt{3}/3] \Rightarrow K(t) = \frac{1}{3!}\left(\frac{(1-t)^4}{4} - \left(\frac{\sqrt{3}}{3}-t\right)^3\right)$

(3) $t\in[\sqrt{3}/3,1] \Rightarrow K(t) = \frac{(1-t)^4}{24}$

Průběh funkce K je zachycen na obrázku 9.2. Z předchozího plyne, že funkce K je integrovatelná a nemění znaménko. Pro výpočet integrálu

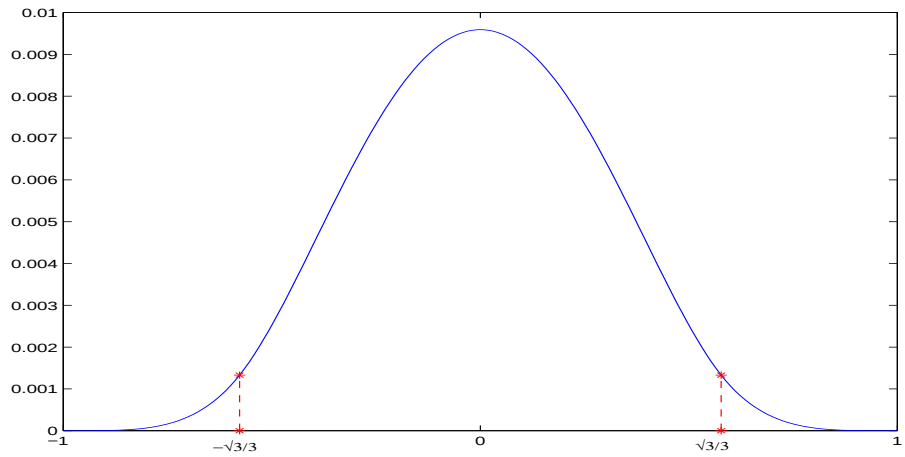
$$R(f) = \int_{-1}^{1} f^{(4)}(t)K(t)\mathrm{d}t$$

lze použít druhé věty o střední hodnotě integrálu:

$$R(f) = f^{(4)}(\alpha)\int_{-1}^{1}K(t)\mathrm{d}t,\ \alpha\in(-1,1).$$

Nyní je možné vypočítat $\int_{-1}^{1}K(t)\mathrm{d}t$, ale protože tento integrál nezávisí na funkci f, lze použít postupu uvedeného v poznámce 5:

$$R(x^4) = \frac{8}{45} \text{ a současně } R(x^4) = 4!\int_{-1}^{1}K(t)\mathrm{d}t.$$



Obr. 9.2: Průběh Peanova jádra.

Odtud

$$\int_{-1}^{1} K(t)\mathrm{d}t = \frac{1}{4!}\frac{8}{45} = \frac{1}{135},$$

takže

$$R(f) = \frac{1}{135} f^{(4)}(\alpha)\,.$$

Příklad 9.5. Užitím Gaussovy-Legendrovy formule (pro $n = 3$) vypočtěte integrál

$$\int_{-1}^{1} \frac{\mathrm{d}x}{1+x^2},$$

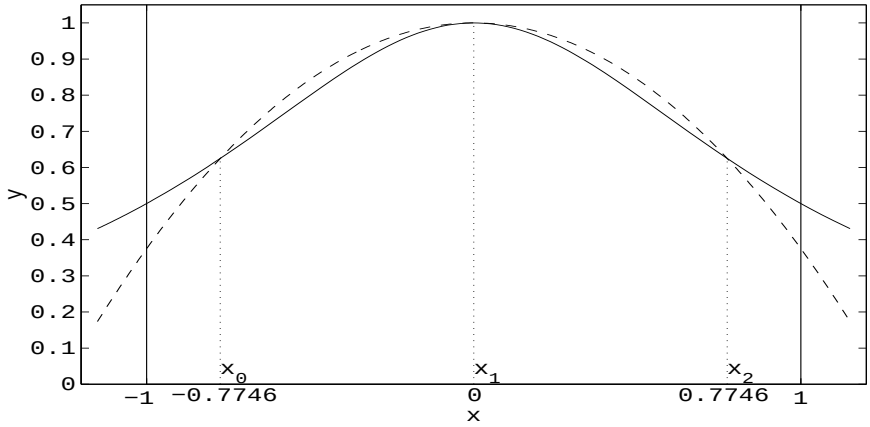
(přesná hodnota je $\frac{\pi}{2}$).
Řešení.

$$\int_{-1}^{1} \frac{\mathrm{d}x}{1+x^2} \approx \frac{2.0{,}651145}{1+(0{,}339981)^2} + \frac{2.0{,}347855}{1+(0{,}861136)^2} \approx 1{,}5668347.$$

V následující tabulce jsou uvedeny uzly, koeficienty a chyba Gaussovy-Legendrovy formule pro $n = 1, 2, 3$.

n	x_i	A_i	$R(f)$
1	$\pm 0{,}577350 = \pm\frac{\sqrt{3}}{3}$	1	$\frac{1}{135} f^{(4)}(\alpha_1)$
2	0	$\frac{8}{9}$	$\frac{1}{15750} f^{(6)}(\alpha_2)$
	$\pm 0{,}774597 = \pm\sqrt{\frac{3}{5}}$	$\frac{5}{9}$	
3	$\pm 0{,}339981$	0,651145	$\frac{1}{3472875} f^{(8)}(\alpha_3)$
	$\pm 0{,}861136$	0,347855	

Obrázky 9.3 a 9.4 ilustrují geometrický význam Gaussovy-Legendrovy formule. Graf funkce je znázorněn plnou čarou a graf polynomu, který integrujeme, je znázorněn čárkovaně.

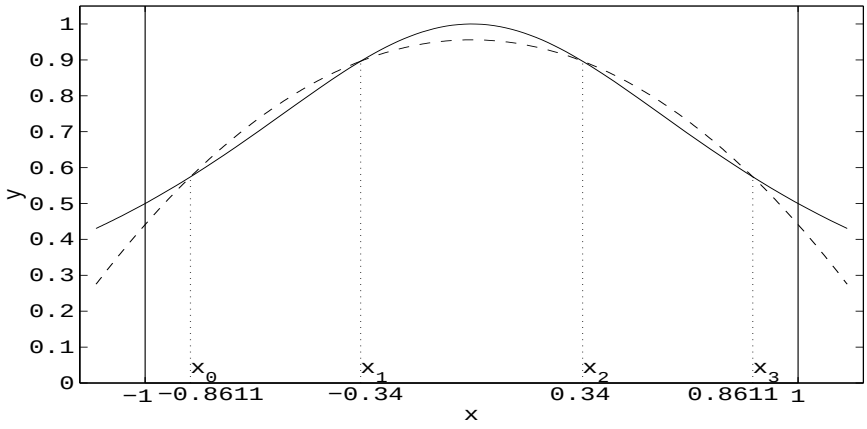


Obr. 9.3: Gaussova-Legendreova kvadraturní formule, $n = 2$

Nechť vahová funkce $w(x) = 1/\sqrt{1-x^2}$, $[a, b] = [-1, 1]$. Polynomy, které jsou ortogonální na tomto intervalu s uvedenou vahovou funkcí, jsou Čebyševovy polynomy. Kvadraturní formule pro výpočet integrálu $\int_{-1}^{1} 1/\sqrt{1-x^2} f(x)\,\mathrm{d}x$ se nazývá *Gaussova-Čebyševova kvadraturní formule*. Tato formule je tvaru

$$\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\,\mathrm{d}x = \sum_{i=0}^{n} A_i f(x_i) + \frac{2\pi}{2^{2(n+1)}(2(n+1)!)} f^{(2n+2)}(\eta), \qquad -1 < \eta < 1.$$

Následující tabulka udává koeficienty, uzly a chybu této formule pro $n = 1, 2, 3$.



Obr. 9.4: Gaussova-Legendreova kvadraturní formule, $n = 3$

n	x_i	A_i	$R(f)$
1	$\pm\frac{\sqrt{2}}{2}$	$\frac{\pi}{2}$	$\frac{\pi}{192}f^{(4)}(\eta_1)$
2	$0, \pm\frac{\sqrt{3}}{2}$	$\frac{\pi}{3}$	$\frac{\pi}{3840}f^{(6)}(\eta_2)$
3	$\pm 0{,}92386$ $\pm 0{,}38268$	$\frac{\pi}{4}$	$\frac{\pi}{5160960}f^{(8)}(\eta_3)$

Vidíme, že u jednotlivých formulí jsou všechny koeficienty stejné a rovné $\pi/(n+1)$.

Příklad 9.6. Užitím Gaussovy-Čebyševovy kvadraturní formule ($n = 2$) vypočtěte integrál

$$\int_{-1}^{1} \frac{\mathrm{d}x}{(1+x^2)\sqrt{1-x^2}}$$

(přesná hodnota je $\pi/\sqrt{2}$).
Řešení. Je

$$\int_{-1}^{1} \frac{\mathrm{d}x}{(x^2+1)\sqrt{1-x^2}} \approx \frac{\pi}{3}\left(\frac{1}{1+\left(-\frac{\sqrt{3}}{2}\right)^2} + \frac{1}{1+0} + \frac{1}{1+\left(\frac{\sqrt{3}}{2}\right)^2}\right) = \frac{5}{7}\pi.$$

Laguerrovy polynomy jsou ortogonální na intervalu $[0, \infty)$ s vahou $w(x) = \mathrm{e}^{-x}$. Kvadraturní formule pro výpočet integrálu $\int_0^\infty \mathrm{e}^{-x} f(x)\,\mathrm{d}x$ se nazývá *Gaussova-*

Laguerrova kvadraturní formule. Formule je tedy tvaru

$$\int_0^\infty e^{-x} f(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i f(x_i) + \frac{((n+1)!)^2}{(2(n+1))!} f^{(2n+2)}(\gamma), \qquad 0 < \gamma < \infty.$$

V následující tabulce jsou uvedeny uzly, koeficienty a chyba této formule pro $n = 1, 2, 3$:

	x_i	A_i	$R(f)$
1	0,585786	0,853553	$\frac{1}{6} f^{(4)}(\gamma_1)$
	3,414214	0,146447	
2	0,415775	0,711093	$\frac{1}{20} f^{(6)}(\gamma_2)$
	2,294280	0,278512	
	6,289945	0,010389	
3	0,322548	0,603154	$\frac{1}{70} f^{(8)}(\gamma_3)$
	1,745761	0,357419	
	4,536620	0,038888	
	9,395071	0,000539	

Příklad 9.7. Užitím Gaussovy-Laguerrovy kvadraturní formule ($n = 1$) vypočtěte

$$\int_0^\infty x^4 \mathrm{e}^{-x}\,\mathrm{d}x,$$

(přesná hodnota je 4!).
Řešení.

$$\int_0^\infty x^4 \mathrm{e}^{-x}\,\mathrm{d}x \approx 0{,}853553.(0{,}585786)^4 + 0{,}146447.(3{,}414214)^4 \approx 20{,}00006.$$

Pro výpočet integrálu $\int_{-\infty}^{\infty} \mathrm{e}^{-x^2} f(x)\,\mathrm{d}x$ lze užít *Gaussovy-Hermitovy* kvadraturní formule. Tyto formule jsou tvaru

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i f(x_i) + \frac{\sqrt{\pi}(n+1)!}{2^{n+1}(2n+2)!} f^{(2n+2)}(\beta), \qquad -\infty < \beta < \infty.$$

Uzly této kvadraturní formule jsou kořeny Hermitova polynomu ortogonálního na intervalu $(-\infty, \infty)$ s vahou $w(x) = \mathrm{e}^{-x^2}$. Následující tabulka uvádí uzly, koeficienty a chybu této formule:

n	x_i	A_i	$R(f)$
1	±0,707107	0,886227	$\frac{\sqrt{\pi}}{48} f^{(4)}(\beta_1)$
2	0	1,181636	$\frac{\sqrt{\pi}}{960} f^{(6)}(\beta_2)$
	±1,224745	0,295409	
3	±0,524648	0,804914	$\frac{\sqrt{\pi}}{26880} f^{(8)}(\beta_3)$
	±1,650680	0,081313	

Příklad 9.8. Užitím Gaussovy-Hermitovy kvadraturní formule ($n = 3$) vypočtěte integrál

$$\int_{-\infty}^{\infty} |x|\, \mathrm{e}^{-x^2}\, \mathrm{d}x,$$

(přesná hodnota je rovna 1).
Řešení.

$$\int_{-\infty}^{\infty} |x|\, \mathrm{e}^{-x^2}\, \mathrm{d}x \approx 2(0{,}524648.0{,}804914 + 1{,}650680.0{,}081313) \approx 1{,}1130364.$$

Poznámka 7. Gaussovy formule mají ještě jednu důležitou vlastnost — s rostoucím počtem uzlů *konverguje posloupnost Gaussových formulí k přesné hodnotě integrálu.* Tuto vlastnost nemají všechny kvadraturní formule, obecně totiž není splněn předpoklad, že součet absolutních hodnot koeficientů je stejnoměrně ohraničený pro všechna n. Pro Gaussovy formule však tento předpoklad splněn je — viz věta 9.5b (podrobněji viz [19], [21]).

Gaussovy formule tvoří velmi důležitou třídu kvadraturních formulí. K jejich zhodnocení a použití se vrátíme v závěru této kapitoly.

Ukážeme nyní na příkladech některé další zajímavé vlastnosti Gaussových formulí.

Uvažujme integrál

$$J = \int_{-1}^{1} w(x) \frac{f(x)}{f(x) + f(-x)}\, \mathrm{d}x, \tag{9.18}$$

kde vahová funkce je sudá a funkce f není lichá. Počítejme tento integrál. Substitucí $y = -x$ vede na integrál

$$J = \int_{-1}^{1} w(y) \frac{f(-y)}{f(y) + f(-y)}\, \mathrm{d}y.$$

Odtud

$$2J = \int_{-1}^{1} w(x) \frac{f(x)}{f(x) + f(-x)}\, \mathrm{d}x + \int_{-1}^{1} w(x) \frac{f(-x)}{f(x) + f(-x)}\, \mathrm{d}x = \int_{-1}^{1} w(x)\, \mathrm{d}x$$

a

$$J = \frac{1}{2}\int_{-1}^{1} w(x)\,\mathrm{d}x. \tag{9.19}$$

Nechť nyní $w(x) \equiv 1$, pak $J = 1$ a vypočtěme tento integrál pomocí Gaussových-Legendrových formulí. Položme

$$F(x) = \frac{f(x)}{f(x) + f(-x)}.$$

Je

$$Q(F) = \sum_{i=0}^{n} A_i F(x_i) = \frac{1}{2}\sum_{i=0}^{n} A_i(F(x_i) + F(-x_i)) = \frac{1}{2}\sum_{i=0}^{n} A_i = 1,$$

tj. $Q(F) = J$. To znamená, že Gaussova-Legendreova formule dává přesnou hodnotu integrálu. Zde jsme použili skutečnosti, že uzly x_i jsou symetricky rozloženy vzhledem k bodu 0 a koeficienty odpovídající symetrickým uzlům jsou stejné (věta 9.3). A dále, víme, že pro koeficienty Gaussových formulí platí

$$\sum_{i=0}^{n} A_i = \int_{-1}^{1} w(x)\,\mathrm{d}x,$$

což znamená v případě Gaussových-Legendrových formulí

$$\sum_{i=0}^{n} A_i = 2.$$

Příklad 9.9.

a) $J = \displaystyle\int_{-1}^{1} \frac{\mathrm{d}x}{1 + \mathrm{e}^{-2x}} = \int_{-1}^{1} \frac{\mathrm{e}^{x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}}\,\mathrm{d}x = 1,$

b) $J = \displaystyle\int_{-1}^{1} \frac{\mathrm{d}x}{\sqrt{1-x^2}(1 + \mathrm{e}^{-2x})} = \int_{-1}^{1} \frac{\mathrm{e}^{x}}{\sqrt{1-x^2}(\mathrm{e}^{x} + \mathrm{e}^{-x})}\,\mathrm{d}x = \frac{\pi}{2}.$

§ 9.3. Newtonovy-Cotesovy kvadraturní formule

Nyní se budeme zabývat kvadraturními formulemi, pro které jsou předepsány všechny uzly. Podrobně vyšetříme případ ekvidistantních uzlů.

Nechť je tedy dáno dělení intervalu $[a, b]$:

$$a = x_0 < x_1 < \ldots < x_n = b,$$

$x_i = x_0 + ih$, $i = 0, 1, \ldots, n$, $h = (b - a)/n$.

Nechť nejdříve $P_n \in \Pi_n$ je Lagrangeův interpolační polynom pro body $(x_i, f(x_i))$, $i = 0, 1, \ldots, n$. Stejným způsobem jako ve větě 9.2 dostaneme formuli

$$\int_a^b w(x)f(x)\,\mathrm{d}x = \sum_{i=0}^{n} A_i f(x_i) + \frac{1}{(n+1)!}\int_a^b w(x)\,\omega_{n+1}(x)f^{(n+1)}(\xi)\,\mathrm{d}x. \quad (9.20)$$

Čísla $A_i = \int_a^b w(x)l_i(x)\,\mathrm{d}x$, $i = 0, 1, \ldots, n$, se nazývají *Cotesova* čísla a jsou tabelována. Uvedená formule se nazývá *Newtonova-Cotesova formule uzavřeného typu*, v tomto případě *integrační meze jsou uzly* kvadraturní formule. Zřejmě je stupeň přesnosti této formule alespoň n. V některých případech lze vyjádření chyby zjednodušit ([1], [8], [19]). Je-li vahová funkce sudá vzhledem ke středu $s = (a+b)/2$ a číslo n je sudé, a protože uzly jsou v tomto případě symetricky rozloženy vzhledem ke středu s, je podle poznámky za větou 9.3 stupeň přesnosti této kvadraturní formule $n+1$. To znamená, že ve vyjádření chyby bude vystupovat $f^{(n+2)}$. Toho lze dosáhnout integrací per partes chybového členu (viz poznámka 4).

Uvedeme nejjednodušší typy těchto formulí ($w(x) \equiv 1$).

a) Nechť $n = 1$. Funkci f aproximujeme polynomem $P_1 \in \Pi_1$, tj.

$$f(x) = \frac{x-a}{b-a}f(b) - \frac{x-b}{b-a}f(a) + \frac{(x-a)(x-b)}{2}f''(\xi).$$

Integrací

$$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{2}(f(a)+f(b)) + \frac{1}{2}\int_a^b (x-a)(x-b)f''(\xi)\,\mathrm{d}x.$$

Spočítejme nyní chybu

$$R(f) = \frac{1}{2}\int_a^b (x-a)(x-b)f''(\xi)\,\mathrm{d}x.$$

Za předpokladu, že f'' je spojitá na intervalu $[a,b]$, lze užít druhé věty o střední hodnotě integrálu, neboť funkce $u(x) = (x-a)(x-b)$ nemění znaménko na intervalu $[a,b]$. Chybu lze nyní vyjádřit takto:

$$R(f) = \frac{f''(\eta)}{2}\int_a^b (x-a)(x-b)\,\mathrm{d}x = -\frac{(b-a)^3}{12}f''(\eta), \qquad a < \eta < b.$$

Výsledná formule je tvaru

$$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{2}(f(a)+f(b)) - \frac{(b-a)^3}{12}f''(\eta). \quad (9.21)$$

Kvadraturní formule $Q(f) = \frac{1}{2}(b-a)(f(a)+f(b))$ je obsah lichoběžníka, a proto se tato formule nazývá *lichoběžníkové pravidlo*.

b) Nechť $n = 2$. V tomto případě aproximujeme f polynomem $P_2 \in \Pi_2$, neboli f nahradíme parabolou. Výsledná formule je tvaru

$$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right) - \frac{1}{90}\left(\frac{b-a}{2}\right)^5 f^{(4)}(\tau) \tag{9.22}$$

Toto pravidlo se nazývá *Simpsonovo* nebo také *parabolické* pravidlo. Důkaz viz cvičení.

Příklad 9.10. Užitím a) lichoběžníkového, b) Simpsonova pravidla vypočtěte integrál

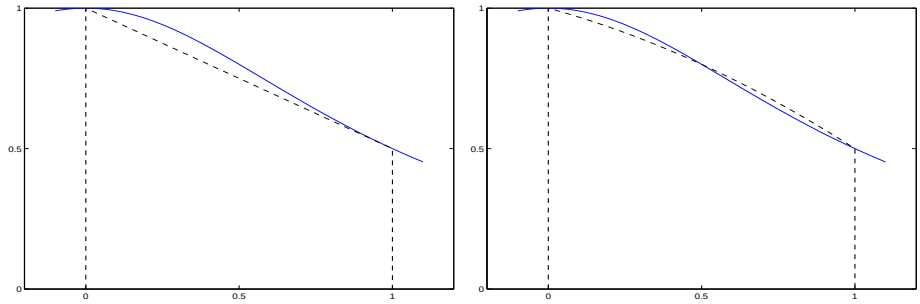
$$\int_0^1 \frac{\mathrm{d}x}{1+x^2}$$

(přesná hodnota je $\frac{\pi}{4}$).

a) $\displaystyle\int_0^1 \frac{\mathrm{d}x}{1+x^2} \approx \frac{1}{2}\left(1 + \frac{1}{2}\right) = \frac{3}{4}$

b) $\displaystyle\int_0^1 \frac{\mathrm{d}x}{1+x^2} \approx \frac{1}{6}\left(1 + 4\frac{4}{5} + \frac{1}{2}\right) = \frac{47}{60}$

Na obrázku 9.5 vidíme průběh integrované funkce a plochy ohraničené čárkovanou křivkou, které se používají k přibližnému výpočtu integrálu lichoběžníkovým a Simpsonovým pravidlem.



Obr. 9.5: Lichoběžníkové a Simpsonovo pravidlo pro funkci $f(x) = \frac{1}{1+x^2}$.

Aproximujme nyní funkci f interpolačním polynomem v bodech $x_1, \ldots, x_{n-1}$, tj. pouze ve „vnitřních“ uzlech intervalu $[a, b]$. Stejným způsobem jako dříve obdržíme formuli

$$\int_a^b f(x)w(x)\,\mathrm{d}x = \sum_{i=1}^{n-1} A_i f(x_i) + \frac{1}{(n-1)!}\int_a^b w(x)\,\omega_{n-1}(x) f^{(n-1)}(\xi)\,\mathrm{d}x,$$

kde A_i, $i = 1, \dots, n$, jsou opět *Cotesova čísla*. Tuto kvadraturní formuli nazýváme *Newtonovou-Cotesovou formulí otevřeného typu*, *integrační meze nejsou uzly* kvadraturní formule. Zřejmě je stupeň přesnosti této formule alespoň $n-2$. Je-li opět vahová funkce sudá vzhledem k $s = (a+b)/2$ a n je číslo sudé, je přesnost formule $n-1$.

Odvodíme nyní nejjednodušší formuli tohoto typu ($w(x) \equiv 1$, $n = 2$). Vahová funkce je v tomto případě sudá a ukážeme, že přesnost této formule je 1.

Funkci f aproximujeme interpolačním polynomem $P_0 \in \Pi_0$ v bodě $((a+b)/2, f((a+b)/2))$, tj.

$$f(x) = f\left(\frac{a+b}{2}\right) + \left(x - \frac{a+b}{2}\right) f'(\xi), \qquad \xi = \xi(x),$$

($P_0(x) \equiv f((a+b)/2)$). Integrujeme:

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x &= \int_a^b f\left(\frac{a+b}{2}\right) \mathrm{d}x + \int_a^b \left(x - \frac{a+b}{2}\right) f'(\xi)\, \mathrm{d}x = \\ &= (b-a) f\left(\frac{a+b}{2}\right) + \int_a^b \left(x - \frac{a+b}{2}\right) f'(\xi)\, \mathrm{d}x.\end{aligned}$$

Vypočteme nyní chybu

$$R(f) = \int_a^b \left(x - \frac{a+b}{2}\right) f'(\xi)\, \mathrm{d}x.$$

Tento integrál budeme počítat metodou per partes a to takto:

$$\begin{aligned} u(x) &= \int_a^x \left(t - \frac{a+b}{2}\right) \mathrm{d}t, & v(x) &= f'(\xi) \\ u'(x) &= \left(x - \frac{a+b}{2}\right), & v'(x) &= \frac{f''(\eta)}{2!} \qquad \eta \in [a,b], \end{aligned}$$

(pro výpočet v' jsme užili věty 7.1). Je zřejmě $u(x) = (x-a)(x-b)/2$ a $u(a) = u(b) = 0$ a $u(x) \le 0$ pro $\forall x \in [a,b]$. Pro náš integrál máme

$$R(f) = [u(x)v(x)]_a^b - \int_a^b u(x)v'(x)\, \mathrm{d}x = -\frac{1}{4}\int_a^b (x-a)(x-b) f''(\eta)\, \mathrm{d}x.$$

Za předpokladu, že $f \in C^2([a,b])$, lze užít druhé věty o střední hodnotě integrálu, neboť funkce u nemění znaménko na tomto intervalu:

$$R(f) = -\frac{f''(\tau)}{4}\int_a^b (x-a)(x-b)\, \mathrm{d}x = \frac{(b-a)^3}{24} f''(\tau_1), \qquad \tau_1 \in (a,b).$$

Výsledná formule

$$\int_a^b f(x)\, \mathrm{d}x = (b-a) f\left(\frac{a+b}{2}\right) + \frac{(b-a)^3}{24} f''(\tau_1)$$

se nazývá *obdélníkové pravidlo*, neboť $Q(f) = (b-a)f((a+b)/2)$ je obsah obdélníka o stranách $(b-a)$ a $f((a+b)/2)$.

Příklad 9.11. Užitím obdélníkového pravidla vypočtěte $\int_0^1 1/(1+x^2)\,\mathrm{d}x$. (přesná hodnota je $\pi/4$.)
Řešení.

$$\int_0^1 \frac{\mathrm{d}x}{1+x^2} \approx 1\frac{1}{1+1/4} = 4/5.$$

Následující tabulky ukazují přehled nejužívanějších Newtonových-Cotesových kvadraturních formulí.

Newtonovy-Cotesovy formule uzavřeného typu:

1. $$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{2}(f(a)+f(b)) - \frac{(b-a)^3}{12}f''(\alpha_1), \qquad a<\alpha_1<b$$

2. $$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{6}\left(f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right) - \left(\frac{b-a}{2}\right)^5 \frac{f^{(4)}(\alpha_2)}{90}, \qquad a<\alpha_2<b$$

3. $$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{8}\left(f(a)+3f\left(a+\frac{b-a}{3}\right)+3f\left(a+2\frac{b-a}{3}\right)+ f(b)\right) - \frac{3}{80}\left(\frac{b-a}{3}\right)^5 f^{(4)}(\alpha_3), \qquad a<\alpha_3<b$$

4. $$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{90}\left(7f(a)+32f\left(a+\frac{b-a}{4}\right)+12f\left(a+\frac{b-a}{2}\right)+ 32f\left(a+3\frac{b-a}{4}\right)+7f(b)\right) - \left(\frac{b-a}{4}\right)^7 \frac{8}{945}f^{(6)}(\alpha_4), \qquad a<\alpha_4<b$$

Formule 3. se také nazývá *Newtonovo pravidlo* nebo *pravidlo 3/8.*

Newtonovy-Cotesovy formule otevřeného typu:

5. $$\int_a^b f(x)\,\mathrm{d}x = (b-a)f\left(\frac{a+b}{2}\right) + \frac{(b-a)^3}{24}f''(\alpha_5), \qquad a<\alpha_5<b$$

6. $$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{2}\left(f\left(a+\frac{b-a}{3}\right)+f\left(a+2\frac{b-a}{3}\right)\right)+ \left(\frac{b-a}{3}\right)^3 \frac{f''(\alpha_6)}{4}, \qquad a<\alpha_6<b$$

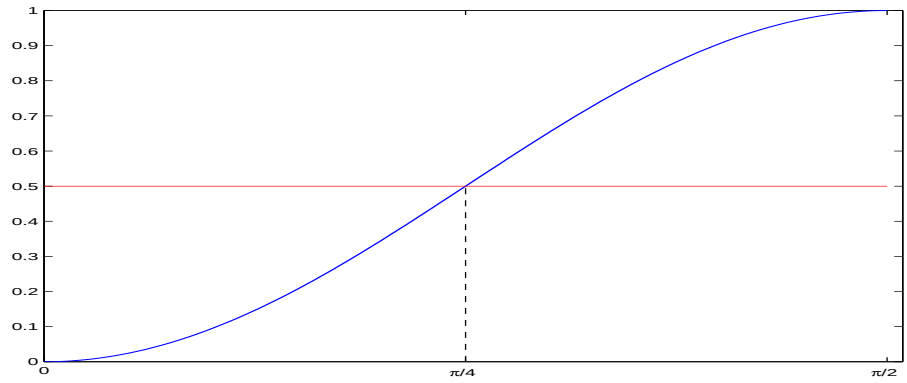
7. $$\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{3}\left(2f\left(a+\frac{b-a}{4}\right)-f\left(a+\frac{b-a}{2}\right)+ 2f\left(a+3\frac{b-a}{4}\right)\right)+\frac{28}{90}\left(\frac{b-a}{4}\right)^5 f^{(4)}(\alpha_7), \qquad a<\alpha_7<b$$

$$8.\ \int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{24}\left(11f\left(a+\frac{b-a}{5}\right)+f\left(a+2\frac{b-a}{5}\right)+\right.$$
$$\left.+f\left(a+3\frac{b-a}{5}\right)+11f\left(a+4\frac{b-a}{5}\right)\right)+$$
$$+\frac{95}{144}\left(\frac{b-a}{5}\right)^5 f^{(4)}(\alpha_8), \qquad a<\alpha_8<b$$

Příklad 9.12. Na závěr tohoto odstavce uvedeme zajímavý příklad. Uvažujme integrál

$$\int_0^{\frac{\pi}{2}} \sin^2 x\mathrm{d}x\,,$$

jehož přesná hodnota je $\pi/4$. Ukážeme, že všechny Newtonovy-Cotesovy formule



Obr. 9.6: Integrace funkce $f(x)=\sin^2 x$.

dávají přesnou hodnotu $\pi/4$:

Víme, že uzly Newtonových-Cotesových formulí jsou ekvidistantní a jsou symetricky rozloženy vzhledem ke středu s intervalu $[a,b]$. V našem případě $s=\pi/4$. Všechny tyto formule mají stupeň přesnosti $N\geq 1$, a tedy zejména pro funkci $f(x)\equiv 1$ platí

$$\int_0^{\frac{\pi}{2}} 1\mathrm{d}x = \sum_{i=0}^{n} A_i \ \Rightarrow\ \sum_{i=0}^{n} A_i = \frac{\pi}{2}\,.$$

Na druhé straně z vlastností funkce $f(x)=\sin^2 x$, $x\in[0,\pi/2]$ plyne opět ze symetrie uzlů vzhledem k bodu $s=\pi/4$, že

$$f(x_i)+f(\pi/2-x_i)=1\,.$$

Zřejmě platí $x_i = i\,\pi/2$, $x_{n-i} = \pi/2 - x_i$, a odtud

$$\sum_{i=0}^{n} A_i f(x_i) = \frac{1}{2}\sum_{i=0}^{n} A_i (f(x_i) + f(\pi/2 - x_i)) = \frac{1}{2}\sum_{i=0}^{n} A_i = \frac{\pi}{4}\,,$$

neboť koeficienty odpovídající symetrickým uzlům jsou stejné.

Podívejme se ještě na výpočet tohoto integrálu pomocí Gaussovy-Legendrovy formule (9.17). Nejdříve je třeba užít substituce tak, abychom daný integrál převedli na integrál s mezemi -1, 1.

$$\int_0^{\frac{\pi}{2}} \sin^2 x \mathrm{d}x = \left| y = \frac{4}{\pi}x - 1 \right| = \frac{\pi}{4}\int_{-1}^{1} \sin^2 \frac{\pi(y+1)}{4} \mathrm{d}y\,.$$

Nyní tato formule dá hodnotu

$$\frac{\pi}{4}\int_{-1}^{1} \sin^2 \frac{\pi(y+1)}{4} \mathrm{d}y = \frac{\pi}{4} 2 \sin^2 \frac{\pi}{4} = \frac{\pi}{4}\,.$$

Dostali jsme opět přesnou hodnotu. Gaussova-Legendreova formule je totiž pro případ $n = 0$ totožná s obdélníkovým pravidlem pro interval $[-1, 1]$!

§ 9.4. Lobattova kvadraturní formule

Zmíníme se stručně o případu, kdy jsou pro kvadraturní formuli předepsány pouze některé uzly. Problematiku objasníme pro integrál

$$I(f) = \int_{-1}^{1} f(x)\,\mathrm{d}x.$$

Naším úkolem je najít kvadraturní formuli tvaru

$$Q(f) = A_0 f(-1) + \sum_{i=1}^{n-1} A_i f(x_i) + A_n f(1).$$

Celkový počet neznámých je $2n$: uzly $x_1, \ldots, x_{n-1}$ a koeficienty $A_0, A_1, \ldots, A_n$. Lze očekávat, že přesnost takové formule bude $2n - 1$, neboť můžeme požadovat splnění následujících podmínek:

$$\int_{-1}^{1} x^k\,\mathrm{d}x = A_0(-1)^k + \sum_{i=1}^{n-1} A_i x_i^k + A_n, \qquad k = 0, 1, \ldots, 2n-1.$$

Nechť nyní $P_1 \in \Pi_1$ je interpolační polynom pro funkci f v bodech -1, 1; je zřejmě

$$P_1(x) = \frac{1-x}{2} f(-1) + \frac{1+x}{2} f(1).$$

Daný integrál vyjádříme ve tvaru

$$I(f) = \int_{-1}^{1} P_1(x)\,\mathrm{d}x + \int_{-1}^{1} \frac{f(x) - P_1(x)}{1 - x^2} w(x)\,\mathrm{d}x, \tag{9.23}$$

kde $w(x) = 1 - x^2$.

Nyní pro výpočet druhého integrálu sestrojíme Gaussovu kvadraturní formuli s vahou $w(x)$:

$$\int_{-1}^{1} \frac{f(x) - P_1(x)}{1 - x^2} w(x)\,\mathrm{d}x \approx \sum_{i=1}^{n-1} \bar{A}_i \frac{f(x_i) - P_1(x_i)}{1 - x_i^2}. \tag{9.24}$$

Zde $\bar{A}_i$ jsou koeficienty Gaussovy kvadraturní formule pro vahovou funkci $w(x) = 1 - x^2$ a $x_1, \ldots, x_{n-1}$ jsou kořeny ortogonálního polynomu s vahou $w(x) = 1 - x^2$ na intervalu $[-1, 1]$.

Připomínáme, že z vlastností ortogonálních polynomů plyne, že $x_i \neq \pm 1$, $i = 1, \ldots, n-1$. Vypočteme-li přímo integrál v (9.23), dostaneme požadovanou kvadraturní formuli

$$\int_{-1}^{1} f(x)\,\mathrm{d}x \approx A_0 f(-1) + \sum_{i=1}^{n-1} A_i f(x_i) + A_n f(1), \tag{9.25}$$

kde

$$\begin{aligned} A_i &= \frac{\bar{A}_i}{1 - x_i^2}, \qquad i = 1, \ldots, n-1, \\ A_0 &= 1 - \sum_{i=1}^{n-1} \bar{A}_i \frac{1}{2(1 + x_i)}, \\ A_n &= 1 - \sum_{i=1}^{n-1} \bar{A}_i \frac{1}{2(1 - x_i)}. \end{aligned}$$

Tato formule se nazývá *Lobattova kvadraturní formule*. Koeficienty a uzly této formule jsou tabelovány a lze je najít např. v [19]. Lze ověřit, že tato formule má přesnost $2n - 1$ a chybu lze vyjádřit ve tvaru ([19]):

$$R(f) = \frac{f^{(2n)}(\tau)}{(2n)!} \int_{-1}^{1} \omega_{2n}(x)\,\mathrm{d}x, \qquad \tau \in (-1, 1),$$

$\omega_{2n}(x) = (x^2 - 1)(x - x_1)^2 \ldots (x - x_{n-1})^2$. (Pro odvození tohoto vztahu lze užít Hermitova interpolačního polynomu.)

Příklad 9.13. Pro $n = 3$ je Lobattova kvadraturní formule tvaru

$$\int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{1}{6}\left(f(-1) + 5f\left(-\frac{\sqrt{5}}{5}\right) + 5f\left(\frac{\sqrt{5}}{5}\right) + f(1)\right) - \\ - \frac{2}{2{,}3625.10^4} f^{(6)}(\eta), \qquad \eta \in (-1, 1).$$

Vypočítáme pomocí této formule integrál:

$$\int_{-1}^{1} \frac{\mathrm{d}x}{1+x^2} \approx \frac{1}{6}\left(\frac{1}{2} + 5\frac{1}{1+\left(-\frac{\sqrt{5}}{5}\right)^2} + 5\frac{1}{1+\left(\frac{\sqrt{5}}{5}\right)^2} + \frac{1}{2}\right) = 1{,}5\overline{5}$$

Příklad 9.14. Pro výpočet integrálu

$$\int_{-1}^{1} f(x)\mathrm{d}x$$

najděte formuli ve tvaru

$$Q(f) = A_0 f(-1) + A_1 f(x_1)\,.$$

Řešení: Lze očekávat stupeň přesnosti $N \geq 2$. Z podmínek $R(x^0) = R(x^1) = R(x^2) = 0$ plyne

$$\begin{aligned} A_0 + A_1 &= 2 \\ A_0(-1) + A_1 x_1 &= 0 \\ A_0(-1)^2 + A_1 x_1^2 &= \frac{2}{3}\,. \end{aligned}$$

Polynom $\omega(x) = (x+1)(x-x_1) = x^2 + a_1 x + a_2$ má kořeny $x_0 = -1$, x_1. Je třeba určit kořen x_1. Jako u odvození Gaussovy formule vynásobíme první rovnici a_2, druhou a_1, třetí jedničkou a sečteme. Výsledkem je rovnice

$$2a_2 + \frac{2}{3} = 0\,.$$

Odtud $a_2 = -1/3$. Polynom $\omega(x) = x^2 + a_1 x - 1/3$ má kořen $x_0 = -1$ tedy $1 - a_1 - 1/3 = 0$ odkud $a_1 = 2/3$ a kořen $x_1 = 1/3$.

Koeficienty A_0, A_1 určíme z prvních dvou rovnic: $A_0 = 1/2$, $A_1 = 3/2$. Výsledná formule je tvaru

$$Q(f) = \frac{1}{2}f(-1) + \frac{3}{2}f(\frac{1}{3})\,.$$

Koeficienty této formule lze rovněž získat integrací interpolačního polynomu v bodech $x_0 = -1$, $x_1 = 1/3$. Odtud pak snadno získáme vyjádření chyby

$$R(f) = \int_{-1}^{1} \frac{(x+1)(x-1/3)}{6} f''(\xi)\mathrm{d}x, \quad \xi = \xi(x).$$

§ 9.5. Čebyševova kvadraturní formule

Podívejme se nyní na kvadraturní formuli z výpočetního hlediska. Formule

$$\int_a^b w(x)f(x)\,\mathrm{d}x \approx \sum_{i=0}^{n} A_i f(x_i)$$

požadují $(n+1)$ násobení a n sčítání. Jestliže však jsou všechny koeficienty stejné, tj. $A = A_i$, $i = 0, 1, \ldots, n$, pak je třeba pouze jedno násobení a n sčítání.

Kvadraturní formule tohoto typu

$$\int_a^b w(x)f(x)\,\mathrm{d}x \approx A\sum_{i=0}^{n} f(x_i) \tag{9.26}$$

se nazývají *Čebyševovy kvadraturní formule*. Tyto formule mají ještě další užitečnou vlastnost, a to:

Jsou-li dány hodnoty $f(x_i)$ s chybami ε_i, pak výsledná chyba způsobená těmito nepřesnostmi bude nejmenší právě v případě, kdy $A = A_i$, $i = 0, 1, \ldots, n$.

V uvedené formuli (9.26) máme k dispozici $(n + 2)$ „volných“ parametrů — koeficient A a uzly $x_0, x_1, \ldots, x_n$. Lze očekávat, že přesnost formule bude $n + 1$; můžeme totiž požadovat splnění rovnic

$$\int_a^b x^k w(x)\,\mathrm{d}x = A\sum_{i=0}^{n} x_i^k, \qquad k = 0, 1, \ldots, n+1. \tag{9.27}$$

Z první rovnice ihned plyne

$$A = \frac{1}{n+1}\int_a^b w(x)\,\mathrm{d}x.$$

Ale řešit soustavu (9.27) není jednoduchá záležitost. Pro interval $[-1, 1]$ a vahovou funkci $w(x) \equiv 1$ řešil tuto úlohu P. L. Čebyšev a nalezl řešení pro $n = 0, 1, 2, 3, 4, 5, 6$. Pro $n = 7$ dokázal, že uzly x_i jsou komplexní čísla.

Tímto problémem se rovněž zabýval S. N. Bernštejn a ukázal, že pro $n > 8$ nemá tato úloha řešení ([2]). Tedy pro interval $[-1, 1]$ a vahovou funkci $w(x) \equiv 1$ lze sestrojit Čebyševovy formule pro $n = 0, 1, 2, 3, 4, 5, 6, 8$ a tyto formule jsou tvaru

$$\int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{2}{n+1}\sum_{i=0}^{n} f(x_i) + R(f).$$

Chybu této formule lze vyjádřit např. užitím Peanovy věty ([8], [5]).

I když Čebyševovy formule uvedeného tvaru lze sestrojit pouze pro malý počet uzlů, je tento počet dostatečný, neboť jak ukážeme dále, užíváme obvykle kvadraturních formulí nižších řádů.

Příklad 9.15. Odvoďte Čebyševovu kvadraturní formuli ve tvaru

$$\int_{-1}^{1} f(x)\mathrm{d}x \approx A(f(x_0) + f(x_1) + f(x_2))\,.$$

Řešení: V tomto případě může být stupeň přesnosti $N \geq 3$. Ze vztahů $R(x^0) = R(x^1) = R(x^2) = R(x^3) = 0$ dostáváme

$$\begin{aligned} A(x_0^0 + x_1^0 + x_2^0) &= 2 \\ A(x_0 + x_1 + x_2) &= 0 \\ A(x_0^2 + x_1^2 + x_2^2) &= \tfrac{2}{3} \\ A(x_0^3 + x_1^3 + x_2^3) &= 0\,. \end{aligned} \tag{9.28}$$

Z první rovnice plyne, že $A = 2/3$. Nechť nyní $\omega_3(x) = (x - x_0)(x - x_1)(x - x_2) = x^3 + a_1x^2 + a_2x + a_3$ je neznámý polynom, jehož kořeny jsou právě uzly x_0, x_1, x_2. Dále první rovnici vynásobíme a_3, druhou a_2, třetí a_1, čtvrtou jedničkou a sečteme. Výsledkem je rovnice

$$2a_3 + \frac{2}{3}a_1 = 0\,.$$

Z Newtonových vztahů mezi kořeny a koeficienty plynomu plyne:

$$x_0 + x_1 + x_2 = -\frac{a_1}{a_0},\ \text{kde } a_0 = 1\,,$$

takže z druhé rovnice systému (9.28) plyne, že $a_1 = 0$. Tedy také $a_3 = 0$. Koeficient a_2 určíme opět použitím Newtonových vztahů:

$$x_0x_1 + x_1x_2 + x_0x_2 = \frac{a_2}{a_0}\,. \tag{9.29}$$

Z výše uvedené soustavy rovnic plyne

$$\begin{aligned} x_0 + x_1 + x_2 &= 0 \\ x_0^2 + x_1^2 + x_2^2 &= 1\,. \end{aligned}$$

Odtud

$$\begin{aligned} (x_0 + x_1 + x_2)^2 - (x_0^2 + x_1^2 + x_2^2) = -1 \quad &\Rightarrow \\ \Rightarrow \quad x_0x_1 + x_1x_2 + x_0x_2 = -\frac{1}{2}\,. \end{aligned}$$

A z (9.29) plyne, že $a_2 = -1/2$.

Polynom ω_3 má koeficienty : $a_0 = 1$, $a_1 = 0$, $a_2 = -1/2$, $a_3 = 0$, tedy $\omega_3 = x^3 - 1/2$ a kořeny tohoto polynomu jsou

$$x_0 = -\frac{\sqrt{2}}{2}, \quad x_1 = 0, \quad x_2 = \frac{\sqrt{2}}{2}.$$

Výsledná formule je tvaru

$$Q(f) = \frac{2}{3}(f(-\frac{\sqrt{2}}{2}) + f(0) + f(\frac{\sqrt{2}}{2})).$$

Chybu této kvadraturní formule lze spočítat integrací chyby při interpolaci polynomem $P_2 \in \Pi_2$ v uzlech $-\sqrt{2}/2$, $\sqrt{2}/2$:

$$R(f) = \frac{1}{3!}\int_{-1}^{1} x(x^2 - \frac{1}{2}) f^{(3)}(\xi(x))\mathrm{d}x = \frac{1}{360} f^{(4)}(\eta)$$

a použít podobného postupu jako při odvození chyby obdélníkového pravidla.

Příklad 9.16. Užitím Čebyševovy formule

$$\int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{2}{3}\left[f\left(-\frac{\sqrt{2}}{2}\right) + f(0) + f\left(\frac{\sqrt{2}}{2}\right)\right] + \frac{1}{360} f^{(4)}(\eta), \quad -1 < \eta < 1,$$

vypočtěte integrál

$$\int_{-1}^{1} \frac{\mathrm{d}x}{1+x^2},$$

jehož přesná hodnota je $\pi/2$.
Řešení.

$$\int_{-1}^{1} \frac{\mathrm{d}x}{1+x^2} \approx \frac{2}{3}\left(\frac{1}{1+\left(-\frac{\sqrt{2}}{2}\right)^2} + 1 + \frac{1}{1+\left(\frac{\sqrt{2}}{2}\right)^2}\right) = \frac{14}{9} = 1{,}5\overline{5}.$$

Pro ilustraci uvedeme ještě další formule Čebyševova typu

$$n = 1: \quad \int_{-1}^{1} f(x)\,\mathrm{d}x = f\left(-\frac{\sqrt{3}}{3}\right) + f\left(\frac{\sqrt{3}}{3}\right) + \frac{1}{135} f^{(4)}(\alpha), \qquad -1 < \alpha < 1,$$

$$n = 3: \quad \int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{1}{2}\left[f\left(-\sqrt{\frac{\sqrt{5}+2}{3\sqrt{5}}}\right) + f\left(-\sqrt{\frac{\sqrt{5}-2}{3\sqrt{5}}}\right) + f\left(\sqrt{\frac{\sqrt{5}-2}{3\sqrt{5}}}\right) + f\left(\sqrt{\frac{\sqrt{5}+2}{3\sqrt{5}}}\right)\right] + \frac{2}{42525} f^{(6)}(\alpha_1),$$

$$-1 < \alpha_1 < 1.$$

Vidíme, že pro $n = 1$ dostáváme formuli totožnou s Gaussovou-Legendrovou kvadraturní formulí. Tento výsledek bylo možné očekávat a je důsledkem toho, že v této formuli jsou oba koeficienty stejné.

§ 9.6. Složené kvadraturní formule

Uvažujme integrál

$$\int_0^\pi \sin x \, \mathrm{d}x,$$

jehož přesná hodnota je 2. Vypočítejme nyní tento integrál lichoběžníkovým pravidlem:

$$Q(f) = \frac{\pi}{2}(\sin \pi + \sin 0) = 0.$$

Vysvětlení spočívá v tom, že lichoběžník v tomto případě degeneruje v úsečku $[0, \pi]$.

Rozdělme nyní interval $[0, \pi]$ na dva subintervaly $[0, \pi/2]$, $[\pi/2, \pi]$ a na každém z nich aplikujme lichoběžníkové pravidlo:

$$\begin{aligned}\int_0^\pi \sin x \, \mathrm{d}x &= \int_0^{\frac{\pi}{2}} \sin x \, \mathrm{d}x + \int_{\frac{\pi}{2}}^\pi \sin x \, \mathrm{d}x \approx \\ &\approx \frac{\pi}{4}\left(\sin 0 + \sin \frac{\pi}{2}\right) + \frac{\pi}{4}\left(\sin \frac{\pi}{2} + \sin \pi\right) = \frac{\pi}{2}.\end{aligned}$$

Při rozdělení intervalu na čtyři subintervaly délky $\pi/4$ dostaneme

$$\int_0^\pi \sin x \, \mathrm{d}x \approx \frac{\sqrt{2}+1}{4}\pi.$$

Uvedený postup lze aplikovat i na další typy kvadraturních formulí. Postupujeme přitom takto:

Daný interval rozdělíme na M subintervalů a na každém z těchto subintervalů aplikujeme kvadraturní formuli $Q(f)$. Tím je dána na intervalu $[a, b]$ nová kvadraturní formule, kterou budeme nazývat *složenou kvadraturní formulí* a budeme psát

$$Q^M(f) = \sum Q(f).$$

Pokud je dělení intervalu $[a, b]$ ekvidistantní s krokem h, budeme rovněž užívat označení $Q_h(f)$, $h = (b - a)/M$.

Nyní se budeme podrobněji zabývat složeným lichoběžníkovým pravidlem, které patří k nejužívanějším formulím tohoto typu.

Nechť je dáno dělení intervalu $[a, b]$:

$$a = x_0 < x_1 < \ldots < x_M = b,$$

přičemž $x_{i+1}-x_i = h$, $i = 0, 1, \ldots, M-1$, $h > 0$. Na každém subintervalu $[x_i, x_{i+1}]$ aproximujeme lichoběžníkovým pravidlem integrál

$$\int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x = \frac{h}{2}(f(x_i) + f(x_{i+1})) - \frac{h^3}{12}f''(\xi_i), \qquad x_i < \xi_i < x_{i+1}.$$

Pro daný integrál dostaneme

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{i=0}^{M-1}\int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x = \frac{h}{2}\sum_{i=0}^{M-1}(f(x_i) + f(x_{i+1})) - \frac{h^3}{12}\sum_{i=0}^{M-1} f''(\xi_i). \tag{9.30}$$

Podívejme se nyní na chybu této aproximace. Je

$$R(f) = -\frac{h^3}{12}\sum_{i=0}^{M-1} f''(\xi_i). \tag{9.31}$$

Předpokládejme, že f'' je spojitá v $[a, b]$ a položme $L = \max_{a\le x\le b} f''(x)$, $l = \min_{a\le x\le b} f''(x)$. Pak

$$Ml \le \sum_{i=0}^{M-1} f''(\xi_i) \le ML$$

a

$$l \le \frac{1}{M}\sum_{i=0}^{M-1} f''(\xi_i) \le L.$$

Nyní, protože f'' je spojitá, musí nabývat každé hodnoty mezi svou maximální a minimální hodnotou na intervalu. Odtud plyne, že existuje takové $\xi \in [a, b]$, že

$$f''(\xi) = \frac{1}{M}\sum_{i=0}^{M-1} f''(\xi_i).$$

Když se vrátíme k vyjádřením (9.30), (9.31), lze zapsat chybu ve tvaru

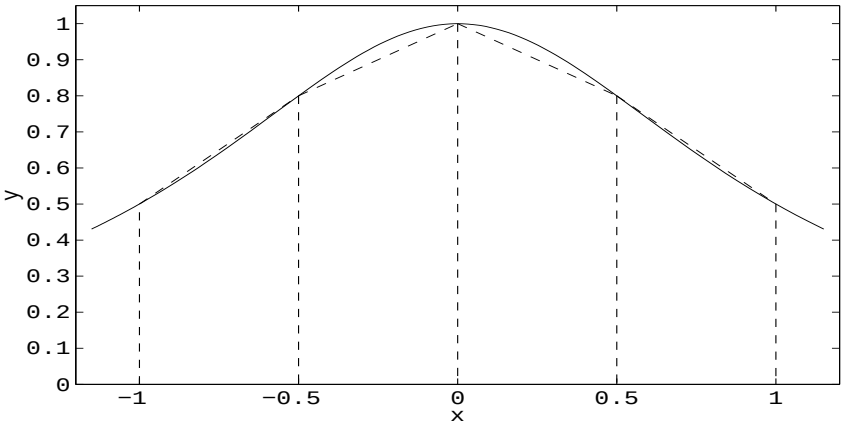
$$R(f) = -\frac{(b-a)^3}{12M^2}f''(\xi)$$

a aproximaci integrálu *složeným lichoběžníkovým pravidlem*

$$\begin{aligned}\int_a^b f(x) = \frac{h}{2}(f(x_0) + 2f(x_1) + \ldots + 2f(x_{M-1}) + f(x_M)) - \\ -\frac{(b-a)^3}{12M^2}f''(\xi), \qquad a\ < \xi < b.\end{aligned} \tag{9.32}$$

Tedy

$$Q^M(f) = \frac{h}{2}(f(x_0) + 2f(x_1) + \ldots + 2f(x_{M-1}) + f(x_M)).$$



Obr. 9.7: Složené lichoběžníkové pravidlo, $f(x) = 1/(1+x^2)$, $Q(f) = 1{,}55$

Poznámka 8. Z tvaru chyby v (9.32) je vidět, že s rostoucím počtem intervalů konvergují aproximace k přesné hodnotě integrálu.

Geometrický význam složeného lichoběžníkového pravidla ilustruje obr. 9.7, kde $M = 4$. Graf dané funkce je vyznačen plnou čarou, odpovídající lichoběžníky jsou vyznačeny čárkovaně.

Při aplikaci složeného lichoběžníkového pravidla je výhodné položit $M_k = 2^k$, $k = 0, 1, 2, \ldots$ a počítat postupně hodnoty formulí $Q^{M_k}(f)$, $k = 0, 1, \ldots$, tj. na každém kroku zdvojnásobit počet subintervalů. Výhodou tohoto postupu je skutečnost, že již jednou vypočtené hodnoty funkce f se použijí ve všech dalších krocích. Tuto vlastnost obecně kvadraturní formule nemají, např. při použití složených Gaussových formulí, je třeba vždy znovu spočítat všechny funkční hodnoty.

Pro toto dělení intervalu můžeme zapsat složené lichoběžníkové pravidlo ve tvaru, který je vhodný zejména pro užití na počítači. Snadno lze totiž ukázat, že

$$Q^{M_k}(f) = \frac{1}{2}\left[Q^{M_{k-1}}(f) + \frac{b-a}{2^{k-1}} \sum_{i=0}^{2^{k-1}} f\left(a + \frac{2^i - 1}{2^k}(b-a)\right)\right], \tag{9.33}$$

kde $M_k = 2^k$, $k = 0, 1, \ldots$

Důkaz tohoto vyjádření ponecháváme do cvičení.

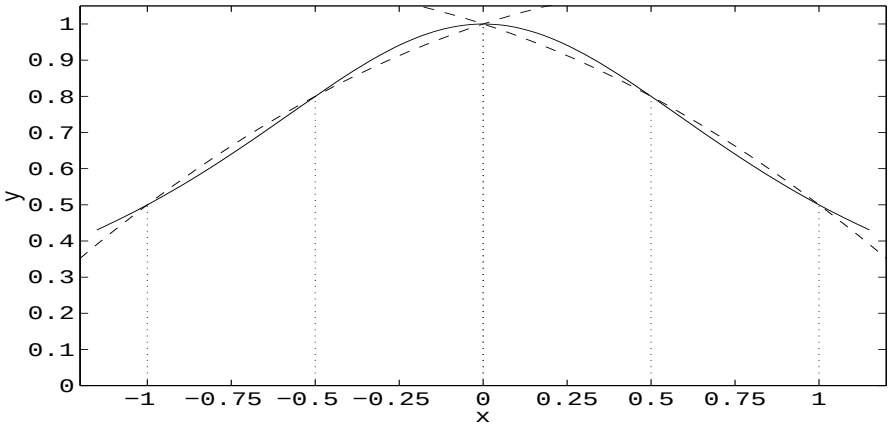
Rozdělme nyní interval $[a, b]$ na $2M$ subintervalů s dělicími body $a = x_0 < x_1 < \ldots < x_{2M} = b$, $x_i - x_{i-1} = h$, $h = (b-a)/2M$. Na každém subintervalu délky $2h$ aplikujme Simpsonovo pravidlo. Výsledná formule

$$\int_a^b f(x)\,\mathrm{d}x = \frac{h}{3}\left[f(a) + 2\sum_{j=1}^{M-1} f(x_{2j}) + 4\sum_{j=1}^{M} f(x_{2j-1}) + f(b)\right] -$$

$$-\frac{b-a}{180}h^4 f^{(4)}(\eta), \qquad a<\eta<b,$$

se nazývá *složené Simpsonovo pravidlo*.

Výraz pro chybu se odvodí stejným způsobem jako u složeného lichoběžníkového pravidla.



Obr. 9.8: Složené Simpsonovo pravidlo, $f(x)=1/(1+x^2)$, $Q(f)=1{,}5667$

Obr. 9.8 ilustruje geometrický význam složeného Simpsonova pravidla pro $M=2$, tj. $h=\frac{1}{2}$, tj. interval $[-1,1]$ rozdělíme na 4 subintervaly délky $h=\frac{1}{2}$ a na 2 subintervalech délky $2h=1$ aplikujeme Simpsonovo pravidlo. Grafy dvou příslušných parabol jsou vyznačeny čárkovaně.

Nyní opět uvedeme zajímavý příklad týkající se aplikace složeného lichoběžníkového pravidla.

Mějme integrál

$$K=\int_0^1 \frac{f(x)}{f(x)+f(1-x)}\,\mathrm{d}x. \tag{9.34}$$

Obdobným způsobem jako při výpočtu integrálu (9.18) lze ukázat, že $K=\frac{1}{2}$. Aplikujme nyní složené lichoběžníkové pravidlo na výpočet tohoto integrálu. Položme

$$F(x)=\frac{f(x)}{f(x)+f(1-x)}.$$

Je jasné, že

$$F(x)+F(1-x)=1. \tag{9.35}$$

Dále víme, že uzly dělení jsou ekvidistantní a $x_{M-i}=1-x_i$, $h=1/M$. Aplikujme nyní složené lichoběžníkové pravidlo na funkci $F(x)$. Je

$$Q^M(F)=\frac{1}{M}\left(\frac{1}{2}F(x_0)+F(x_1)+\ldots+F(x_{M-1})+\frac{1}{2}F(x_M)\right)=$$

$$= \frac{1}{2M}\left(\frac{1}{2}F(x_0) + F(x_1) + \ldots + F(x_{M-1}) + \frac{1}{2}F(x_M) + \right.$$
$$\left. + \frac{1}{2}F(1-x_0) + F(1-x_1) + \ldots + F(1-x_{M-1}) + \frac{1}{2}F(x_0)\right)$$

Vezmeme-li v úvahu (9.35), dostaneme

$$Q^M(F) = \frac{1}{2M}\left(\frac{1}{2} + (M-1) + \frac{1}{2}\right) = \frac{1}{2}.$$

Vidíme, že složené lichoběžníkové pravidlo dává přesnou hodnotu integrálu K.

§ 9.7. Adaptivní kvadraturní formule

Jak jsme již uvedli, složené kvadraturní formule obvykle užívají ekvidistantního dělení intervalu. Ale tento postup není vhodný v případě, kdy integrační interval obsahuje jak subintervaly, kde funkce značně osciluje, tak také subintervaly, kde funkční hodnoty se nemění příliš rychle. To znamená, že v prvním případě je vhodný menší krok dělení, aby příslušná kvadraturní formule vhodně vystihla chování funkce, v druhém případě krok dělení může být větší. Efektivní metody, které mohou přizpůsobit délku kroku variaci funkce, se nazývají *adaptivní kvadraturní formule*.

V systému MATLAB jsou uvedeny dvě metody tohoto typu:

1) QUAD — adaptivní kvadraturní formule založená na Simpsonově pravidle,
2) QUAD 8 — adaptivní kvadraturní formule založená na pravidle 3/8.

Vysvětlíme stručně podstatu metody QUAD. Je třeba aproximovat integrál

$$I(f) = \int_a^b f(x)\,\mathrm{d}x$$

s chybou menší než ε, $\varepsilon > 0$. Označme $S(a,b)$ Simpsonovo pravidlo pro interval $[a,b]$,

$$S(a,b) = \frac{h}{3}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right), \qquad h = \frac{b-a}{2}.$$

Pak

$$\int_a^b f(x)\,\mathrm{d}x = S(a,b) - \frac{h^5}{90}f^{(4)}(\tau), \qquad a < \tau < b. \tag{9.36}$$

Aplikujme nyní složené Simpsonovo pravidlo pro dva subintervaly $[a,(a+b)/2]$, $[(a+b)/2,b]$, tj. krok dělení je $h_1 = h/2 = (b-a)/4$. Aproximace integrálu je nyní tvaru

$$\int_a^b f(x)\,\mathrm{d}x = \frac{h}{6}\left(f(a) + 4f\left(a+\frac{h}{2}\right) + 2f(a+h) + 4f\left(a+\frac{3}{2}h\right) + f(b)\right) -$$
$$- \left(\frac{h}{2}\right)^4 \frac{b-a}{180}f^{(4)}(\tilde{\tau}), \qquad a < \tilde{\tau} < b. \tag{9.37}$$

Položme

$$S\left(a, \frac{a+b}{2}\right) = \frac{h}{6}\left(f(a) + 4f\left(a + \frac{h}{2}\right) + f(a+h)\right),$$
$$S\left(\frac{a+b}{2}, b\right) = \frac{h}{6}\left(f(a+h) + 4f\left(a + \frac{3}{2}h\right) + f(b)\right).$$

Vztah (9.37) můžeme nyní zapsat ve tvaru

$$\int_a^b f(x)\,\mathrm{d}x = S\left(a, \frac{a+b}{2}\right) + S\left(\frac{a+b}{2}, b\right) - \frac{1}{16}\frac{h^5}{90}f^{(4)}(\tilde{\tau}). \tag{9.38}$$

Předpokládejme nyní, že $f^{(4)}$ se příliš nemění v $[a, b]$, tj. předpokládejme, že $f^{(4)}(\tau) \approx f^{(4)}(\tilde{\tau})$. Za tohoto předpokladu vztahy (9.36) a (9.38) implikují

$$S\left(a, \frac{a+b}{2}\right) + S\left(\frac{a+b}{2}, b\right) - \frac{1}{16}\frac{h^5}{90}f^{(4)}(\tau) \approx S(a,b) - \frac{h^5}{90}f^{(4)}(\tau).$$

Odtud

$$\frac{h^5}{90}f^{(4)}(\tau) \approx \frac{16}{15}\left(S(a,b) - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right).$$

Užijeme-li tohoto vyjádření ve vztahu (9.38), dostaneme

$$\begin{aligned}\left|\int_a^b f(x)\,\mathrm{d}x - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right| \approx \\ \approx \frac{1}{15}\left|S(a,b) - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right|.\end{aligned} \tag{9.39}$$

Tento vztah znamená, že součet $(S(a,(a+b)/2) + S((a+b)/2, b))$ aproximuje daný integrál 15krát lépe než (tentýž součet) aproximuje hodnotu $S(a,b)$.

To znamená, že $S(a,(a+b)/2) + S((a+b)/2, b)$ bude aproximovat integrál $I(f)$ s přesností ε, za předpokladu, že $S(a,(a+b)/2) + S((a+b)/2, b)$ se liší od $S(a,b)$ o méně než 15ε, tj. je-li

$$\left|S(a,b) - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right| < 15\varepsilon, \tag{9.40}$$

pak

$$\left|I(f) - S\left(a, \frac{a+b}{2}\right) - S\left(\frac{a+b}{2}, b\right)\right| < \varepsilon. \tag{9.41}$$

Je-li tedy splněna podmínka (9.40), je integrál aproximován s dostatečnou přesností.

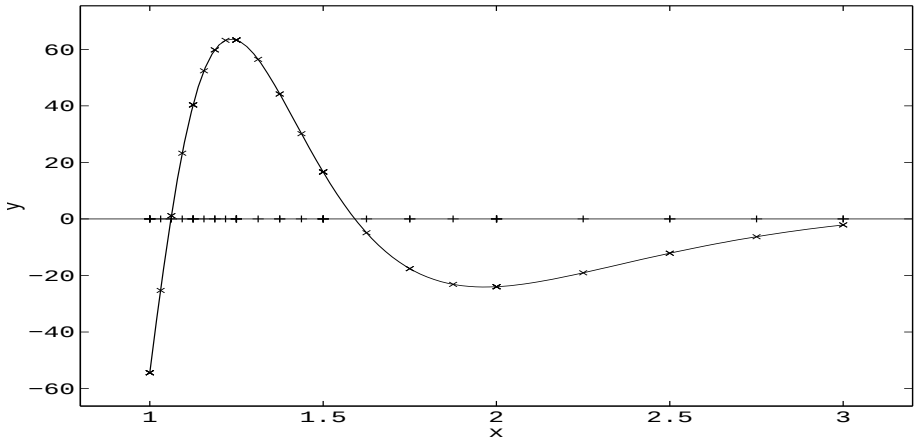
Jaká je tedy základní myšlenka programu QUAD? Testujeme, zda je splněna podmínka (9.40). Je-li splněna, je integrál aproximován s dostatečnou přesností. Jestliže podmínka (9.40) není splněna, aplikujeme výše uvedený postup na každý

subinterval $[a,(a+b)/2]$, $[(a+b)/2,b]$ zvlášť a požadujeme, aby chyba byla na každém z těchto subintervalů menší než $\varepsilon/2$. Pak je celková chyba menší než ε. Je-li chyba aproximace na některém ze subintervalů větší než $\varepsilon/2$, pak se opět rozdělí tento subinterval na dva intervaly a požaduje se, aby chyba aproximace na každém z nich byla menší než $\varepsilon/4$. Postup opakujeme tak dlouho, pokud není dosaženo požadované přesnosti.

Obr. 9.9 ilustruje použití adaptivní formule QUAD. Na ose x jsou vyznačeny koncové body intervalů, které odpovídají dělení při realizaci procedury QUAD,

$$f(x) = \frac{100}{x^2} \sin \frac{10}{x},$$

tolerance 0,03, $Q(f) = -1{,}413$.



Obr. 9.9: QUAD pro výpočet integrálu $\int_1^3 \frac{100}{x^2} \sin \frac{10}{x} \mathrm{d}x$

Program QUAD 8 je konstruován podobným způsobem založeným na pravidle 3/8.

§ 9.8. Rombergova integrace

Lze ukázat ([5], [19]), že složené lichoběžníkové pravidlo s krokem h můžeme vyjádřit ve tvaru

$$Q_h(f) = \int_a^b f(x)\,\mathrm{d}x + c_1 h^2 + c_2 h^4 + \ldots + c_m h^{2m} + \alpha_{m+1}(h) h^{2m+2}, \qquad (9.42)$$

kde $|\alpha_{m+1}(h)| \leq A_{m+1}$ pro všechna $h = (b-a)/M$, $M = 1, 2, \ldots$

Na tomto vyjádření je založena velice užitečná a elegantní *Rombergova kvadraturní formule*. Poslední člen ve vyjádření (9.42) je relativně malý a můžeme jej tedy zanedbat,

$$Q_h(f) \approx I(f) + \sum_{i=1}^{m} c_i h^{2i}, \qquad I(f) = \int_a^b f(x)\,\mathrm{d}x. \tag{9.43}$$

Položme

$$P_m(y) = I(f) + c_1 y + \ldots + c_m y^m. \tag{9.44}$$

Je zřejmě $P_m \in \Pi_m$, $P_m(0) = I(f)$. Naším úkolem je najít přibližnou hodnotu $P_m(0)$. Budeme postupovat takto:

Z (9.43) a (9.44) plyne přibližný vztah

$$P_m(h^2) \approx Q_h(f). \tag{9.45}$$

Uvažujme nyní posloupnost $\{\Delta_k\}$ dělení intervalu $[a,b]$; Δ_k: $x_i = x_0 + ih_k$, $i = 0,1,\ldots,k$, $x_0 = a$, $x_k = b$, $h_k = (b-a)/2^k$, $k = 0,1,\ldots,m$. Nechť $Q_{h_k}(f)$ je hodnota lichoběžníkového pravidla odpovídající kroku h_k. Podle (9.45) je

$$P_m(h_k^2) \approx Q_{h_k}(f), \qquad k = 0,1,\ldots,m. \tag{9.46}$$

Pro funkci $P_m(y)$ známe (přibližné) hodnoty v $(m+1)$ různých bodech $y_k = h_k^2 = ((b-a)/2^k)^2$. Sestrojíme pro tyto hodnoty $(y_k, P_m(y_k))$, $k = 0,\ldots,m$, příslušný Lagrangeův interpolační polynom:

$$P_m(y) = \sum_{k=0}^{m} l_k(y) P_m(y_k) \approx \sum_{k=0}^{m} l_k(y) Q_{h_k}(f),$$

kde l_k, $k = 0,\ldots,m$, jsou fundamentální polynomy (rovnost plyne z jednoznačnosti interpolačního polynomu). Spočítejme hodnotu $P_m(0)$:

$$P_m(0) \approx \sum_{k=0}^{m} l_k(0) Q_{h_k}(f) = \sum_{k=0}^{m} d_k Q_{h_k}(f),$$

kde jsme položili $l_k(0) = d_k$, $k = 0,1,\ldots,m$. Pro výpočet daného integrálu jsme nalezli formuli

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{k=0}^{m} d_k Q_{h_k}(f),$$

kterou nazýváme *Rombergovou kvadraturní formulí*.

Rombergova formule patří mezi tzv. extrapolační metody, neboť hodnotu $P_m(0)$ jsme získali *extrapolací* (bod 0 leží vně intervalu $[h_m^2, h_0^2]$).

Poznámka 9. Výpočet integrálu Rombergovou formulí lze velmi efektivně provést užitím Nevillova schematu pro iterovanou interpolaci.

Položme nejdříve $Q_{h_k}(f) = T_{k0}$, $k = 0, \ldots, m$. Výpočet lze uspořádat do následující tabulky (stejně jako u Nevillova schematu):

$$\begin{array}{lllll} h_0^2 & T_{00} & & & \\ h_1^2 & T_{10} & T_{11} & & \\ h_2^2 & T_{20} & T_{21} & T_{22} & \\ h_3^2 & T_{30} & T_{31} & T_{32} & T_{33} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

kde (viz Nevillovo schema, kap. 6)

$$T_{ij}(y) = \frac{(y - y_{i-j})T_{i,j-1}(y) - (y - y_i)T_{i-1,j-1}(y)}{y_i - y_{i-j}}, \qquad \begin{array}{l} j = 1, 2, 3, \ldots \\ i = j, j+1, \ldots \end{array}$$

a pro $y = 0$ a $y_i = ((b-a)/2^i)^2$ dostaneme rekurentní vztah

$$\begin{aligned} T_{ij} &= \frac{4^j T_{i,j-1} - T_{i-1,j-1}}{4^j - 1}, \qquad \begin{array}{l} j = 1, 2, \ldots, m \\ i = j, j+1, \ldots, m \end{array} \\ T_{mm} &= P_m(0) \approx I(f). \end{aligned} \tag{9.47}$$

Každý člen tabulky T_{ij} představuje ve skutečnosti lineární kvadraturní formuli pro krok délky $h_j = (b-a)/2^j$:

$$T_{ij} = \alpha_0 f(a) + \alpha_1 f(a + h_j) + \ldots + \alpha_{j-1} f(b - h_j) + \alpha_j f(b).$$

Některé z těchto formulí, ale ne všechny, jsou pro $i = k$ formule Newtonova typu, např. T_{11} je Simpsonovo pravidlo:

$$T_{11} = \tfrac{4}{3}T_{10} - \tfrac{1}{3}T_{00} = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right).$$

Poznámka 10. Lze ukázat ([4]), že pro chybu Rombergovy integrace platí

$$T_{ij} - \int_a^b f(x)\,\mathrm{d}x = (b-a)\left(\frac{(b-a)^2}{2^{2i-j}}\right)^{j+1} \frac{(-1)^j B_{2j+2}}{(2j+2)!} f^{(2j+2)}(\xi), \qquad a < \xi < b,$$

pro funkci $f \in C^{2j+2}([a,b])$, B_{2j+2} jsou Bernoulliova čísla.

Příklad 9.17. Užijte Rombergovy kvadraturní formule pro výpočet $\int_0^\pi \sin x\,\mathrm{d}x$.
Řešení. Položme $h_0 = \pi$, $h_1 = \frac{\pi}{2}$, $h_2 = \frac{\pi}{4}$, $h_3 = \frac{\pi}{8}$. Příslušná tabulka hodnot T_{ij}:

h_i^2	T_{i0}	T_{i1}	T_{i2}	T_{i3}
h_0^2	0			
h_1^2	1,57079630	2,09439511		
h_2^2	1,86911890	2,00455976	1,99857073	
h_3^2	1,97423160	2,00026917	1,99998313	2,000005

Lze ukázat, že posloupnost na diagonále v takové tabulce bude konvergovat k hodnotě integrálu rychleji než posloupnost $\{T_{i0}\}$.

Rombergova metoda má ještě jednu důležitou vlastnost: umožňuje snadno přidat další řádek — znamená to pouze vypočítat další lichoběžníkové pravidlo a užít vzorce (9.47). Výpočet lze totiž uspořádat tak, aby proběhl po řádcích, tj. v pořadí T_{00}, T_{10}, T_{11}, T_{20}, T_{21}, $T_{22}, \ldots$ Jako testu pro zastavení výpočtu lze například použít vztahu $|T_{mm} - T_{m-1,m-1}| < \varepsilon$, $\varepsilon > 0$ je předepsaná tolerance, nebo $|T_{mm} - T_{m,m-1}| < \varepsilon$.

Rombergova metoda patří k velmi často používaným metodám. Je výhodná především v těch výpočtech, kde je požadována malá chyba aproximace. I když je tato metoda velmi efektní, není ji možné používat univerzálně. Je to vhodná metoda pro hladké funkce a je vždy třeba předpokládat existenci dostatečného počtu derivací funkce f na celém intervalu $[a, b]$. V opačném případě nepřináší Rombergova metoda žádné zrychlení konvergence oproti např. složenému lichoběžníkovému pravidlu.

Zmíníme se ještě stručně o kvadraturních formulích s rychle oscilujícími vahovými funkcemi, které se například vyskytují při výpočtu Fourierových koeficientů. Aplikace známých kvadraturních formulí na integrály tohoto typu nedává dobré výsledky. Příčina selhání těchto metod je v rychlé oscilaci integrandu, o jehož průběhu nemají použité metody dostatek informace. Výpočtem takových integrálů se zabýval J. Mikloško ([15]).

§ 9.9. Metoda polovičního kroku, použití kvadraturních formulí

Budeme se nyní zabývat otázkou volby kvadraturní formule. Uvedeme pouze faktory, které mohou mít vliv na tuto volbu, ale obecná jenoznačná pravidla stanovit nelze.

V případě, že funkce je dána tabulkou, tj. jsou známy její hodnoty na ekvidistantní množině uzlů, je vhodné použít Newtonových-Cotesových formulí. Jestliže můžeme odhadnout derivaci dané funkce, je možné stanovit z odhadu chyby kvadraturní formule počet uzlů potřebných k dosažení požadované přesnosti. Tyto hodnoty se však v praxi obtížně odhadují a navíc Newtonovy-Cotesovy formule vysokých řádů nekonvergují k přesné hodnotě integrálu (viz [16], [19] a příklad 9.18). Proto se zpravidla postupuje tak, že užijeme složené kvadraturní formule s nižším počtem uzlů a postupně zvyšujeme počet intervalů tak dlouho, až se počet desetinných míst odpovídajících žádané přesnosti stabilizuje ve dvou po sobě jdoucích aproximacích.

Runge navrhl poněkud preciznější způsob odhadu dosažené přesnosti (tzv. metodu *polovičního kroku*). Popíšeme nyní tuto metodu. Z úvah v § 9.6 je vidět, že výpočet pomocí složeného lichoběžníkového resp. Simpsonova pravidla může být zapsán ve tvaru

$$I(f) = Q^M(f) + Kh^2$$

resp.

$$I(f) = Q^M(f) + \tilde{K}\tilde{h}^4,$$

kde K resp. $\tilde{K}$ je součin konstanty a druhé resp. čtvrté derivace f v jistých bodech intervalu $[a, b]$. Dá se ukázat i obecně ([1]), že výpočet pomocí složené kvadraturní formule může být zapsán ve tvaru

$$I(f) \approx Q^M(f) + Kh^N, \qquad h = \frac{b-a}{M},$$

výraz $h^N K$ se nazývá *hlavní člen chyby* a $N-1$ je stupeň přesnosti kvadraturní formule, K závisí na N-té derivaci funkce f. Výpočet veličiny K je zpravidla dosti obtížný. A proto se v praxi postupuje jiným způsobem. Zdvojnásobíme-li počet subintervalů, dostaneme aproximaci

$$I(f) \approx Q^{2M} + K_1 \left(\frac{h}{2}\right)^N.$$

Pokud se derivace funkce f v intervalu $[a, b]$ příliš nemění, lze položit $K \approx K_1$. Platí tedy

$$I(f) \approx Q^M(f) + Kh^N,$$
$$I(f) \approx Q^{2M}(f) + K\left(\frac{h}{2}\right)^N.$$

Z těchto dvou vztahů vypočteme konstantu K:

$$K \approx 2^N \frac{Q^{2M}(f) - Q^M(f)}{(2h)^N - h^N}$$

Nyní pro další aproximaci integrálu obdržíme

$$I(f) \approx Q^{2M}(f) + \frac{Q^{2M}(f) - Q^M(f)}{2^N - 1}.$$

Veličinu

$$z = \frac{Q^{2M}(f) - Q^M(f)}{2^N - 1}$$

lze užít pro odhad chyby. Při výpočtu postupujeme takto: Počítáme složené kvadraturní formule Q^{M_k}, $M_k = 2^k M_0$, $M_0 > 0$, $k = 0, 1, 2, \ldots$ a na každém kroku počítáme veličinu

$$z^k = \frac{Q^{M_{k+1}}(f) - Q^{M_k}(f)}{2^N - 1}.$$

Jestliže pro nějaké $k = l$ je $|z^l| < \varepsilon$, ε je požadovaná přesnost, pak za novou aproximaci integrálu lze vzít hodnotu

$$I(f) \approx Q^{M_{l+1}}(f) + z^l,$$

přičemž chyba této aproximace $|z^l| < \varepsilon$. V podstatě se jedná o aplikaci Richardsonovy extrapolace. Pro $N = 2$ je to právě případ uvedený v kapitole 7.

Je-li funkce dána analyticky, je třeba vzít v úvahu jak Newtonovy-Cotesovy vzorce, tak i Gaussovu kvadraturní formuli. Gaussovy formule dosahují vyšší přesnosti při užití menšího počtu uzlů než Newtonovy-Cotesovy formule a potřebují tedy počítat méně funkčních hodnot funkce f. Mají rovněž příznivější chybový výraz. Požadované přesnosti lze dosáhnout volbou Gaussova vzorce dostatečně vysokého řádu, neboť posloupnost Gaussových formulí konverguje k přesné hodnotě integrálu. Toto jsou výhody Gaussovy formule v případě, že používáme jednu formuli. Při použití složených formulí je však aplikace Gaussových formulí náročnější oproti Newtonovým-Cotesovým formulím. Totiž, při dělení daného intervalu na $M_k = 2^k M_0$, $M_0 =$ přirozené číslo, $k = 0, 1, 2, \ldots$, intervalů lze již jednou vypočtených funkčních hodnot užít ve všech dalších aplikacích Newtonových-Cotesových formulí, ale není tomu tak u Gaussových formulí. A proto použití složených Gaussových formulí je náročnější. Pokud jde o složená pravidla, lze doporučit pro hladké funkce s výhodou složené lichoběžníkové pravidlo a Rombergovu integraci. V případě, že integrand v určité oblasti daného intervalu značně osciluje, ale v jiné oblasti má hladký průběh, je vhodné použít adaptivních formulí.

Na závěr tohoto odstavce si ukážeme, že Newtonovy-Cotesovy formule obecně nekonvergují (viz [19]).

Příklad 9.18. Počítejme přibližnou hodnotu integrálu

$$\int_{-4}^{4} \frac{\mathrm{d}x}{1+x^2} = 2 \arctan 4 \doteq 2,6516353\,.$$

Následující tabulka přibližnou hodnotu integrálu určenou pomocí Newtonovy-Cotesovy formule pro daný počet uzlů a pro porovnání je uveden výsledek získaný složeným lichoběžníkovým pravidlem.

n	Newtonova-Cotesova formule	složené lich. pravidlo
3	5,490	4,235
5	2,278	2,918
7	3,329	2,701
9	1,941	2,659
11	3,596	2,6511
13	1,335	2,6505

Pro $M_7 = 2^7$, tj. 129 funkčních hodnot by pomocí složeného lichoběžníkova pravidla vyšel integrál přibližně $2,651617$, zatímco Newtonova-Cotesova formule dává hodnotu 2.977249 a pro $n = 131$ hodnotu $1,556509$.

§ 9.10. Integrály se singularitami

V předchozích odstavcích jsme se zabývali výpočtem integrálu

$$\int_a^b f(x)\,\mathrm{d}x$$

a předpokládali jsme, že funkce f má dostatečný počet derivací v intervalu $[a, b]$. Ale v praxi se často setkáváme s případy, že integrál nebo jeho derivace mají singularity v $[a, b]$. V tomto odstavci navrhneme několik způsobů, jak postupovat v takových případech.

Omezíme se na singularity v krajních bodech intervalu, neboť rozdělením intervalu na subintervaly lze „zvládnout“ i singularity uvnitř intervalu.

Při výpočtu takových integrálů se ukazuje velmi výhodný postup s vahovými funkcemi. Daný integrand f lze totiž zapsat ve tvaru

$$f(x) = w(x)g(x),$$

kde w je vahová funkce zahrnující singularitu, g je dostatečně hladká funkce. Pro standardní vahové funkce lze užít příslušných Gaussových kvadraturních formulí.

Jako příklad uvedeme formuli pro výpočet integrálu

$$\int_0^1 \frac{g(x)}{\sqrt{x}}\,\mathrm{d}x;$$

zde máme singularitu v levém krajním bodě a vahová funkce je $w(x) = 1/\sqrt{x}$. Určíme ortogonální polynomy s vahou $1/\sqrt{x}$ na intervalu $[0, 1]$. Víme, že pro Legendrovy polynomy platí:

$$\int_0^1 P_m(x)P_n(x)\,\mathrm{d}x = 0 \qquad \text{pro } m \neq n.$$

Dále, Legendrův polynom sudého stupně obsahuje pouze sudé mocniny x a je tedy sudou funkcí. Předchozí vztah pak můžeme zapsat ve tvaru

$$0 = \int_{-1}^1 P_{2m}(x)P_{2n}(x)\,\mathrm{d}x = 2\int_0^1 P_{2m}(x)P_{2n}(x)\,\mathrm{d}x, \qquad m \neq n.$$

Užijeme-li nyní substituce $x^2 = y$, dostaneme

$$\int_0^1 \frac{1}{\sqrt{y}} P_{2n}(\sqrt{y})P_{2m}(\sqrt{y})\,\mathrm{d}y = 0.$$

Odtud, polynomy $p_n(x) = P_{2n}(\sqrt{x})$, $n = 0, 1, 2, \ldots$, jsou ortogonální s vahou $1/\sqrt{x}$ na intervalu $[0, 1]$. Nyní již můžeme sestrojit kvadraturní formuli

$$\int_0^1 \frac{g(x)}{\sqrt{x}}\,\mathrm{d}x \approx \sum_{j=0}^{n} H_j g(a_j),$$

kde

$$a_j = x_j^2, \qquad j = 0, \ldots, n,$$

x_j je kladný kořen polynomu $P_{2(n+1)}$ (polynom $P_{2(n+1)}$ má kořeny $\pm x_0, \ldots, \pm x_n$, $x_i \neq 0$, $\forall i = 0, \ldots, n$). Vypočteme nyní koeficienty H_j, $j = 0, \ldots, n$:

Nechť A_j je koeficient Gaussovy-Legendrovy formule odpovídající kladnému kořenu x_j, $j = 0, 1, \ldots, n$. Je

$$\begin{aligned} A_j &= \int_{-1}^{1} \frac{(x^2-x_0^2)\ldots(x+x_j)(x^2-x_{j+1}^2)\ldots(x^2-x_m^2)}{(x_j^2-x_0^2)\ldots(2x_j)(x_j^2-x_{j+1}^2)\ldots(x_j^2-x_m^2)}\mathrm{d}x = \\ &= \int_{-1}^{1} \frac{x}{2x_j}\frac{(x^2-x_0^2)\ldots(x^2-x_{j-1}^2)(x^2-x_{j+1}^2)\ldots(x^2-x_m^2)}{(x_j^2-x_0^2)\ldots(x_j^2-x_{j-1}^2)(x_j^2-x_{j+1}^2)\ldots(x_j^2-x_m^2)}\mathrm{d}x + \\ &+ \int_{-1}^{1} \frac{x_j}{2x_j}\frac{(x^2-x_0^2)\ldots(x^2-x_{j-1}^2)(x^2-x_{j+1}^2)\ldots(x^2-x_m^2)}{(x_j^2-x_0^2)\ldots(x_j^2-x_{j-1}^2)(x_j^2-x_{j+1}^2)\ldots(x_j^2-x_m^2)}\mathrm{d}x\,. \end{aligned}$$

První integrál je roven nule, neboť integrand je lichá funkce. Pro druhý integrál užijeme substituci $x^2 = y$:

$$\begin{aligned} A_j &= \frac{1}{2}\int_{-1}^{1} 2\frac{(y-x_0^2)\ldots(y-x_{j-1}^2)(y-x_{j+1}^2)\ldots(y-x_m^2)}{(x_j^2-x_0^2)\ldots(x_j^2-x_{j-1}^2)(x_j^2-x_{j+1}^2)\ldots(x_j^2-x_m^2)}\frac{\mathrm{d}y}{2\sqrt{y}} = \\ &= \frac{1}{2}\int_{-1}^{1} \frac{(y-x_0^2)\ldots(y-x_{j-1}^2)(y-x_{j+1}^2)\ldots(y-x_m^2)}{(x_j^2-x_0^2)\ldots(x_j^2-x_{j-1}^2)(x_j^2-x_{j+1}^2)\ldots(x_j^2-x_m^2)}\frac{\mathrm{d}y}{\sqrt{y}} = \\ &= \frac{1}{2}H_j\,. \end{aligned}$$

Pro koeficienty H_j tedy platí

$$H_j = 2A_j, \quad j = 0, \ldots, n\,,$$

kde A_j je koeficient odpovídající kladnému kořenu x_j v Gaussově-Legendrově kvadraturní formuli.

Obdobně postupujeme v případě vahové funkce $w(x) = \sqrt{x}$, $x \in [0,1]$, při výpočtu integrálu

$$\int_0^1 \sqrt{x} g(x)\,\mathrm{d}x.$$

Integrand má tentokrát singularitu v derivaci. Polynomy ortogonální na intervalu $[0,1]$ s vahou $\sqrt{x}$ jsou tvaru

$$p_n(x) = \frac{1}{\sqrt{x}} P_{2n+1}(\sqrt{x}),$$

kde P_{2n+1} je opět Legendrův polynom stupně $2n+1$ (podrobněji viz [19]).

Je-li třeba vypočítat integrál

$$\int_0^1 \left(\frac{x}{1-x}\right)^{\frac{1}{2}} g(x)\,\mathrm{d}x,$$

zvolíme za vahovou funkci $w(x) = (x/(1-x))^{1/2}$. Polynomy ortogonální s vahou $(x/(1-x))^{1/2}$ na intervalu $[0,1]$ jsou určeny vztahem

$$p_n(x) = \frac{1}{\sqrt{x}} T_{2n+1}(\sqrt{x}),$$

kde T_{2n+1} je Čebyševův polynom stupně $2n+1$.

Pro výpočet nevlastních integrálů

$$\int_0^\infty f(x)\,\mathrm{d}x, \qquad \int_{-\infty}^\infty f(x)\,\mathrm{d}x$$

můžeme užít přímo Laguerrovy nebo Hermitovy kvadraturní formule.

Nevlastní integrál

$$\int_a^\infty f(x)\,\mathrm{d}x, \qquad a > 0,$$

jestliže existuje, můžeme rovněž aproximovat užitím vhodné kvadraturní formule a to tak, že jej substitucí $t = x^{-1}$ nejdříve převedeme na integrál

$$\int_0^1 \frac{1}{t^2} f\left(\frac{1}{t}\right)\,\mathrm{d}t.$$

Některé další způsoby výpočtu singulárních integrálů lze najít např. v [8], [19].

Cvičení ke kapitole 9

1. Určete koeficienty A_0, A_1, A_2 tak, aby přesnost kvadraturní formule

$$\int_{-1}^1 f(x)\,\mathrm{d}x = A_0 f\left(-\tfrac{1}{2}\right) + A_1 f(0) + A_2 f\left(\tfrac{1}{2}\right) + R(f)$$

byla alespoň 2.
($A_0 = \frac{4}{3}$, $A_1 = -\frac{2}{3}$, $A_2 = \frac{4}{3}$.)

2. Určete koeficienty A_0, A_1 a uzel x_0 pro formuli

$$\int_0^1 \sqrt{x} f(x)\,\mathrm{d}x = A_0 f(x_0) + A_1 f(1) + R(f).$$

($A_0 = \frac{7}{15}$, $A_1 = \frac{1}{5}$, $x_0 = \frac{3}{7}$.)

3. Určete algebraicky neznámé uzly x_0, x_1 a koeficienty A_0, A_1 pro formuli

$$\int_0^\pi \sin x f(x)\,\mathrm{d}x = A_0 f(x_0) + A_1 f(x_1) + R(f)$$

tak, aby bylo dosaženo maximálního stupně přesnosti.

($A_0 = A_1 = 1$, $x_{0,1} = \dfrac{\pi}{2} \pm \sqrt{\dfrac{\pi^2}{4} - 2}$.)

4. Odvoďte Newtonovu-Cotesovu formuli otevřeného typu pro interval $[-2, 3]$ s krokem $h = 1$.
($\int_{-2}^{3} f(x)\,\mathrm{d}x = \frac{5}{24}(11f(-1) + f(0) + f(1) + 11f(2)) + \frac{95}{144}f^{(4)}(\eta)$, $-2 < \eta < 3$.)

5. Odvoďte Newtonovu-Cotesovu formuli uzavřeného typu pro interval $[a, b]$ a $n = 3$ (tzv. pravidlo 3/8).
($\int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{8}\left(f(a) + 3f\left(a + \frac{b-a}{3}\right) + 3f\left(a + \frac{b-a}{3}\right) + f(b)\right) - \frac{3}{80}\left(\frac{b-a}{3}\right)^5 f^{(4)}(\eta), a < \eta < b$.)

6. Odvoďte Simpsonovo pravidlo.

7. Nechť $f \in C^{(6)}([-1, 1])$ a nechť $P_5 \in \Pi_5$ je Hermitův interpolační polynom s vlastnostmi $P(x_i) = f(x_i)$, $P'(x_i) = f'(x_i)$, $x_i = -1, 0, 1$.

   a) Ukažte, že

   $$\int_{-1}^{1} P(x)\,\mathrm{d}x = \tfrac{7}{15}f(-1) + \tfrac{16}{15}f(0) + \tfrac{7}{15}f(+1) + \tfrac{1}{15}f'(-1) - \tfrac{1}{15}f'(+1).$$

   b) Formule v části a) představuje kvadraturní formuli přesnou pro polynomy stupně nejvýše 5. Ukažte, že formule není přesná pro polynomy stupně 6.

8. Odvoďte formuli Čebyševova typu ve tvaru

   $$\int_{-1}^{1} f(x)\,\mathrm{d}x = A(f(x_0) + f(x_1) + f(x_2)) + R(f).$$

   ($A = \frac{2}{3}$, $x_0 = -\frac{\sqrt{2}}{2}$, $x_1 = 0$, $x_2 = \frac{\sqrt{2}}{2}$, $R = \dfrac{f^{(4)}(\eta)}{360}$, $-1 < \eta < 1$.)

9. Aproximujte integrál

   $$\int_0^{\frac{\pi}{4}} \sin x\,\mathrm{d}x = 1 - \frac{\sqrt{2}}{2}$$

   a) obdélníkovým, b) lichoběžníkovým, c) Simpsonovým pravidlem.
   (a) 0,30055887, b) 0,27768018, c) 0,29293264.)

10. Následující integrály vypočtěte a) lichoběžníkovým, b) Simpsonovým pravidlem. Výsledky porovnejte s přesnými hodnotami

    $$1.\ \int_1^2 \ln x\,\mathrm{d}x, \quad 2.\ \int_0^{0,1} x^{\frac{1}{3}}\,\mathrm{d}x, \quad 3.\ \int_0^{\frac{\pi}{3}} (\sin x)^2\,\mathrm{d}x.$$

    (a) 1. 0,34657, 2. 0,023208, 3. 0,39270,
    b) 1. 0,38583, 2. 0,032296, 3. 0,30543.)

11. Užijte Newtonovy-Cotesovy formule uzavřeného typu pro $n = 3$ (viz cv. 5) pro výpočet

$$\int_1^3 \mathrm{e}^{-\frac{x}{2}}\,\mathrm{d}x.$$

(0,766801.)

12. Užijte a) složeného lichoběžníkového, b) složeného Simpsonova pravidla pro výpočet integrálů:

1. $\int_0^3 x\sqrt{1+x^2}\,\mathrm{d}x, \quad M = 6,$
2. $\int_0^1 \sin \pi x\,\mathrm{d}x, \qquad M = 6,$
3. $\int_0^{2\pi} x \sin x\,\mathrm{d}x, \qquad M = 8,$
4. $\int_0^1 x^2\mathrm{e}^x\,\mathrm{d}x, \qquad M = 8.$

Porovnejte získané aproximace s přesnými hodnotami.
(a) 1. 10,3122, 2. 0,62201, 3. −5,9568, 4. 0,72889,
b) 1. 10,20751, 2. 0,6366357, 3. −6,284027, 4. 0,7182830.)

13. Užijte Rombergovy integrace pro výpočet hodnoty $T_{4,4}$ pro následující integrály a výsledky porovnejte s přesnými hodnotami.

$$\text{a)} \quad \int_0^{\frac{\pi}{4}} \sin x\,\mathrm{d}x \qquad\qquad \text{b)} \quad \int_{\frac{\pi}{2}}^{\frac{3\pi}{4}} \cos x\,\mathrm{d}x$$

14. Užijte Rombergovy metody integrace pro výpočet

$$\int_0^2 x^2\mathrm{e}^{-x^2}\,\mathrm{d}x.$$

Číslo m určete během výpočtu tak, aby $|T_{m,m-1} - T_{m,m}| < 10^{-6}$.
(0,4227250.)

15. Dokažte vztah (9.33).

16. Užitím Gaussovy-Legendrovy kvadraturní formule pro $n = 2, 3, 4$ aproximujte integrál

$$\int_1^3 \mathrm{e}^x \sin x\,\mathrm{d}x.$$

(11,141495; 10,948403; 10,950140.)

17. Opakujte cvičení 10 užitím Gaussových-Legendrových formulí pro $n = 1$.

18. Užitím Gaussovy-Laguerrovy formule pro $n = 2, 3$ aproximujte integrál

$$\int_0^\infty \mathrm{e}^{-x} \sin x \, \mathrm{d}x.$$

($n = 2 \ldots$ 0,432460, $n = 3 \ldots$ 0,496023, přesná hodnota je 0,5.)

19. Odvoďte složené obdélníkové pravidlo a navrhněte příslušný algoritmus.

20. Navrhněte algoritmus pro výpočet

$$\int_a^b f(x) \, \mathrm{d}x$$

se zadanou přesností ε, který vychází z poměrně hrubého dělení intervalu $[a, b]$ a užívá metody poloviční kroku.

21. Pomocí Gaussovy-Čebyševovy formule pro $n = 1, 2$ aproximujte integrál

$$\int_{-1}^1 \frac{\cos x}{\sqrt{1 - x^2}} \, \mathrm{d}x.$$

(2,3884; 2,4041; přesná hodnota 2,40394.)

Kontrolní otázky ke kapitole 9

1. Dávají všechny Gaussovy-Legendreovy formule pro integrál

$$\int_0^{\frac{\pi}{2}} \sin^2 x \mathrm{d}x$$

stejnou hodnotu $\pi/4$ jako Newtonovy-Cotesovy formule?

2. Budou dávat Newtonovy-Cotesovy formule také přesnou hodnotu integrálu (9.18)?

3. Jaká je hodnota integrálu (9.18), je-li w lichá funkce?

4. Jaký tvar má složené obdélníkové pravidlo?

5. Souvisí Rombergova integrace s Richardsonovou extrapolací?

Kapitola 10

Metoda nejmenších čtverců

Při interpolační aproximaci jsme požadovali, aby interpolační polynom nabýval v daných bodech týchž hodnot jako aproximovaná funkce. Ale v případě, že hodnoty funkce jsou dány empiricky a jsou zatíženy „šumem“, není tento přístup nejvhodnější. V této kapitole se budeme zabývat postupem, jímž se z funkčních hodnot zatížených nepřesnostmi sestrojuje taková aproximace, která „vyrovnává“ empirické hodnoty v tom smyslu, že informace o přesném průběhu funkce obsažená v naměřených hodnotách se zachová, ale „šum“ se tímto „vyrovnáním“ odstraní.

Formulujme nyní tuto úlohu přesněji:
Nechť jsou dány body x_j, $j = 0, \ldots, N$, $x_j \neq x_k$, pro $j \neq k$, reálná funkce f, dále reálné funkce ϕ_i, $i = 0, \ldots, m$, a hodnoty těchto funkcí v bodech x_j, $j = 0, \ldots, N$. Je třeba najít koeficienty $c_0, \ldots, c_m$ lineární kombinace[1]

$$P_m(x) = c_0\phi_0(x) + \ldots + c_m\phi_m(x) \tag{10.1}$$

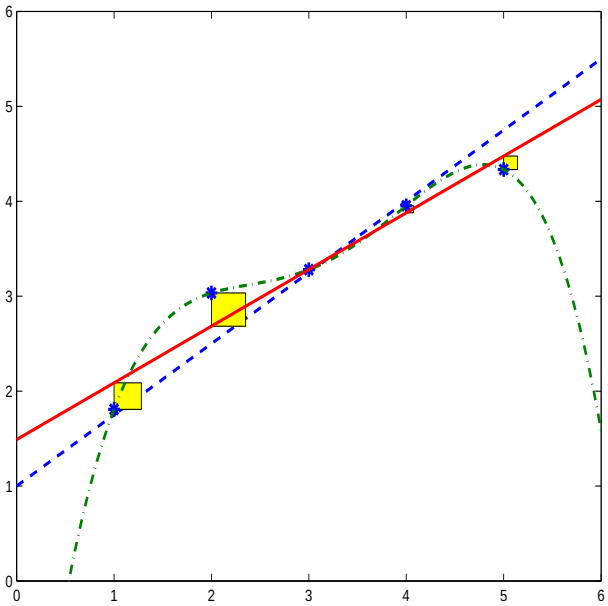
tak, aby funkce

$$\rho^2(c_0, \ldots, c_m) = \sum_{j=0}^{N} (f_j - P_m(x_j))^2, \quad f_j = f(x_j) \tag{10.2}$$

nabývala minimální hodnoty, tj. minimalizujeme součet čtverců. Odtud plyne také název *metoda nejmenších čtverců*. Funkci P_m, pro kterou ρ^2 nabývá minimální hodnoty, nazýváme *nejlepší aproximaci funkce f na množině* $\{x_j\}_{j=0}^N$.

Poznámka 1. Budeme se zabývat případem $m < N$, neboť pro $m \geq N$ je možné řešit úlohu interpolací.
Geometrický význam metody je demonstrován na obrázku 10.1. Nejlepší aproximace je taková, pro kterou součet obsahů jednotlivých čtverců o stranách $|f_j - P_m(x_j)|$, $j = 0, \ldots, N$, je minimální.

[1] Omezíme se zde pouze na lineární případ, otázky nelineární aproximace jsou studovány např. v [5].



Obr. 10.1: Demonstrace metody nejmenších čtverců:

– – – –	původní funkce
* * * * *	hodnoty s „šumem“
– · – · –	interpolační polynom
———	odhad metodou nejmenších čtverců
□	obsah $(f_j - P_m(x_j))^2$, $j = 0, \ldots, N$

Úmluva. Nechť

$$(f, g) = \sum_{j=0}^{N} f(x_j) f(x_j) \tag{10.3}$$

označuje skalární součin funkcí f a g na množině $\{x_j\}_{j=0}^{N}$.

Funkce $\rho^2(c_0, \ldots, c_m)$ je funkce $m+1$ proměnných. Při hledání minima postupujeme jako v analýze: derivujme (10.2) podle každé z neznámých c_k, $k = 0, \ldots, m$, a derivace položme rovny nule, tj.

$$\frac{\partial \rho^2}{\partial c_k} = -2 \sum_{j=0}^{N} \left(f_j - \sum_{i=0}^{m} c_i \phi_i(x_j) \right) \phi_k(x_j) = 0, \quad k = 0, \ldots, m.$$

Užitím (10.3) lze tyto rovnice zapsat ve tvaru

$$\begin{array}{ccccccccc} c_0(\phi_0,\phi_0) & + & c_1(\phi_0,\phi_1) & + & \dots & + & c_m(\phi_0,\phi_m) & = & (f,\phi_0) \\ c_0(\phi_1,\phi_0) & + & c_1(\phi_1,\phi_1) & + & \dots & + & c_m(\phi_1,\phi_m) & = & (f,\phi_1) \\ & & & & & & & \vdots & \\ c_0(\phi_m,\phi_0) & + & c_1(\phi_m,\phi_1) & + & \dots & + & c_m(\phi_m,\phi_m) & = & (f,\phi_m) \end{array} \tag{10.4}$$

Vektorový zápis soustavy lze odvodit následujícím způsobem. Položme

$$\boldsymbol{f} = \begin{pmatrix} f_0 \\ \vdots \\ f_N \end{pmatrix}, \qquad \boldsymbol{P} = \begin{pmatrix} P_0 \\ \vdots \\ P_N \end{pmatrix}, \qquad P_i = P(x_i),\ i = 0\ldots,N$$

$$A = \begin{pmatrix} \phi_0(x_0) & \dots & \phi_m(x_0) \\ \vdots & & \vdots \\ \phi_0(x_N) & \dots & \phi_m(x_N) \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} c_0 \\ \vdots \\ c_N \end{pmatrix}.$$

Pak $\boldsymbol{P} = A\boldsymbol{c}$ a

$$\rho^2(\boldsymbol{c}) = (\boldsymbol{f}-\boldsymbol{P})^T(\boldsymbol{f}-\boldsymbol{P}) = (\boldsymbol{f}-A\boldsymbol{c})^T(\boldsymbol{f}-A\boldsymbol{c})\,.$$

Systém (10.4) je potom možné přepsat ve tvaru

$$G\boldsymbol{c} = \boldsymbol{d}\,,$$

kde

$$G = A^TA = \begin{pmatrix} (\phi_0,\phi_0) & \dots & (\phi_0,\phi_m) \\ \vdots & & \vdots \\ (\phi_m,\phi_0) & \dots & (\phi_m,\phi_m) \end{pmatrix}, \qquad \boldsymbol{d} = A^T\boldsymbol{f} = \begin{pmatrix} (f,\phi_0) \\ \vdots \\ (f,\phi_m) \end{pmatrix},$$

takže soustavu (10.4) lze také psát ve tvaru

$$A^TA\,\boldsymbol{c} = A^T\boldsymbol{f}\,. \tag{10.5}$$

Nyní se budeme zabývat otázkami řešitelnosti soustavy (10.4).

Definice 10.1. Soustava lineárních rovnic (10.4) se nazývá *normální.* soustava. Její determinant se nazývá Gramův determinant příslušný funkcím $\phi_0,\dots,\phi_m$.

Definice 10.2. Řekneme, že funkce $\phi_0,\dots,\phi_m$ jsou lineárně nezávislé na množině bodů $x_0,\dots,x_N$, jestliže jsou lineárně nezávislé vektory $(\phi_0(x_0),\dots,\phi_0(x_N))^T$, $\dots$, $(\phi_m(x_0),\dots,\phi_m(x_N))^T$.

Z definice plyne, že funkce $\phi_0,\dots,\phi_m$ jsou lineárně závislé na množině bodů $x_0,\dots,x_N$, jestliže existují čísla $a_0,\ \dots,\ a_m$, $|a_0|+\ldots+|a_m|\neq 0$, tak, že platí

$$a_0\phi_0(x_j)+\ldots+a_m\phi_m(x_j) = 0, \quad j = 0,\dots,N.$$

Platí následující tvrzení

Věta 10.1. *Gramův determinant je různý od nuly právě tehdy, když funkce ϕ_0, ..., ϕ_m jsou lineárně nezávislé na množině bodů $x_0, \ldots, x_N$.*

Důkaz. Matice $G = A^T A$ je regulární právě tehdy, když má hodnost $m+1$. Hodnost matice G je ovšem rovna hodnosti matice A, jejíž sloupce tvoří vektory $(\phi_0(x_0), \ldots, \phi_0(x_N))^T$, ..., $(\phi_m(x_0), \ldots, \phi_m(x_N))^T$. Ty jsou lineárně nezávislé v případě, že funkce $\phi_0, \ldots, \phi_m$ jsou lineárně nezávislé na množině bodů $x_0, \ldots, x_N$. □

Poznámka 2. Ze vztahu $G = A^T A$ navíc bezprostředně plyne, že G je pozitivně semidefinitní, respektive pozitivně definitní (v případě regularity).

Věta 10.2. *Nechť funkce $\phi_0, \ldots, \phi_m$ jsou lineárně nezávislé na množině bodů $x_0, \ldots, x_N$. Pak normální soustava (10.4) má jediné řešení $c_0^*, \ldots, c_m^*$ a funkce*

$$P_m(x) = c_0^* \phi_0(x) + \ldots + c_m^* \phi_m(x)$$

je nejlepší aproximací funkce f na množině $\{x_j\}_{j=0}^N$.

Důkaz. Funkce $\phi_0, \ldots, \phi_m$ jsou lineárně nezávislé na množině bodů $x_0, \ldots, x_N$. Odtud plyne, že determinant soustavy, je různý od nuly a soustava má jediné řešení $c_0^*, \ldots, c_m^*$. Ukážeme, že pro toto řešení nabývá funkce ρ^2 minimální hodnoty. Položme $\boldsymbol{c}^* = (c_0^*, \ldots, c_m^*)$ a počítejme veličinu $\rho^2(\boldsymbol{c}^* + \Delta\boldsymbol{c})$, kde $\Delta\boldsymbol{c} \neq \boldsymbol{o}$.

$$\begin{aligned}
\rho^2(\boldsymbol{c}^* + \Delta\boldsymbol{c}) &= (\boldsymbol{f} - A\,\boldsymbol{c}^* - A\,\Delta\boldsymbol{c})^T(\boldsymbol{f} - A\,\boldsymbol{c}^* - A\,\Delta\boldsymbol{c}) = \\
&= (\boldsymbol{f} - A\,\boldsymbol{c}^*)^T(\boldsymbol{f} - A\,\boldsymbol{c}^*) - \Delta\boldsymbol{c}^T A^T(\boldsymbol{f} - A\,\boldsymbol{c}^*) - \\
&- (\boldsymbol{f} - A\,\boldsymbol{c}^*)^T A\,\Delta\boldsymbol{c} + \Delta\boldsymbol{c}^T A^T A\,\Delta\boldsymbol{c} = \\
&= \rho^2(\boldsymbol{c}^*) - \Delta\boldsymbol{c}^T(A^T\boldsymbol{f} - A^T A\,\boldsymbol{c}^*) - (A^T\boldsymbol{f} - A^T A\,\boldsymbol{c}^*)^T\Delta\boldsymbol{c} + \\
&+ \Delta\boldsymbol{c}^T A^T A\,\Delta\boldsymbol{c}
\end{aligned}$$

Člen $(A^T\boldsymbol{f} - A^T A\,\boldsymbol{c}^*)$ roven nulovému vektoru, protože $\boldsymbol{c}^*$ je řešení soustavy (10.5). Navíc je výraz $\Delta\boldsymbol{c}^T A^T A\,\Delta\boldsymbol{c}$ kladný, neboť matice $A^T A$ je pozitivně definitní a $\Delta\boldsymbol{c}$ je nenulový vektor. Celkem tedy $\rho^2(\boldsymbol{c}^* + \Delta\boldsymbol{c}) > \rho^2(\boldsymbol{c}^*)$. □

Důsledek. *Pro funkce $\phi_i(x) = x^i$, $i = 0, \ldots, m$, $m < N$ má normální soustava jediné řešení.*

Příklad 10.1. Užijte metody nejmenších čtverců k nalezení nejlepší lineární aproximace pro hodnoty

x_i	-1	1	3	5	7
f_i	1	3	4	5	6

Řešení. Nejlepší aproximaci budeme hledat ve tvaru

$$P_1(x) = c_0 + c_1 x\,.$$

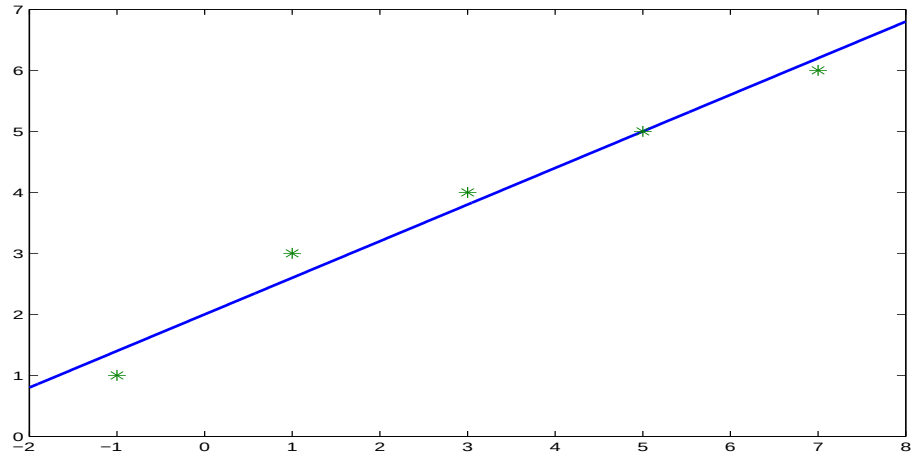
Veličina ρ^2 je v tomto případě tvaru

$$\begin{aligned}\rho^2(c_0,c_1) &= (1-c_0+c_1)^2+(3-c_0-c_1)^2+(4-c_0-3c_1)^2+\\ &+ (5-c_0-5c_1)^2+(6-c_0-7c_1)^2\end{aligned}$$

a normální soustava

$$\begin{aligned} 5c_0 &+ 15c_1 &= 19\\ 15c_0 &+ 85c_1 &= 81\,, \end{aligned}$$

jejíž řešení je $c_0^* = 2$, $c_1^* = \frac{3}{5}$. Hledaná nejlepší lineární aproximace je $P_1(x) = 2+\frac{3}{5}x$ – viz obr. 10.2



Obr. 10.2: Výsledky příkladu 1.

Poznámka 3. Pro lineárně nezávislé funkce ϕ_i, $i = 0,\dots,m$ je matice normální soustavy symetrická a pozitivně definitní. Tyto soustavy lze řešit Gaussovou eliminační metodou nebo Choleského metodou (viz kapitola 4).

Zmíníme se podrobněji o volbě $\phi_i(x) = x^i$, $i = 0,\dots,m$. V tomto případě při větším stupni aproximujícího polynomu ($m > 5$) je soustava (10.4) špatně podmíněná, tzn. že řešení je značně citlivé na vliv chyb ve vstupních datech a na zaokrouhlovacích chybách během výpočtu. Dá se totiž ukázat (viz [19]), že determinant soustavy je přibližně $(N+1)$ násobkem matice

$$H = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{m+1} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{m+2} \\ \vdots & \vdots & & \vdots \\ \frac{1}{m+1} & \frac{1}{m+2} & \cdots & \frac{1}{2m+1} \end{pmatrix},$$

která je klasickým příkladem špatně podmíněné matice (inverzní matice má velmi velké prvky – pro $m = 9$ prvky řádu 10^{12}). Důsledkem špatné podmíněnosti je

skutečnost, že při řešení soustavy způsobí každá zaokrouhlovací chyba, které se dopustíme, mnohonásobně větší chybu v řešení. Již pro $m = 9$ se mohou vyskytovat při řešení značné obtíže.

Dalším problémem je volba hodnoty m, tj. stupně polynomu (viz podrobněji [19], [18]).

Zvolíme-li však za funkce ϕ_i, $i = 0, \ldots, m$. diskrétně *ortogonální polynomy*, tj. platí-li pro funkce ϕ_i

$$(\phi_i, \phi_k) = 0 \quad \text{pro } i \neq k,$$

tyto problémy odpadají. Matice normální soustavy je v tomto případě diagonální a soustava je tvaru

$$\begin{array}{rrrcl} c_0(\phi_0,\phi_0) & & & = & (f,\phi_0) \\ & c_1(\phi_1,\phi_1) & & = & (f,\phi_1) \\ & & \ddots & \vdots & \\ & & c_m(\phi_m,\phi_m) & = & (f,\phi_m) \end{array}$$

Odtud lze snadno vypočítat koeficienty nejlepší aproximace a

$$P_m(x) = \sum_{i=0}^{m} \frac{(f,\phi_i)}{(\phi_i,\phi_i)} \phi_i(x). \tag{10.6}$$

Pro metodu nejmenších čtverců lze s výhodou užít diskrétních Čebyševových nebo Gramových polynomů.

a) *Čebyševovy polynomy* (diskrétní případ). Uvažujme skalární součin

$$(f,g) = \sum_{j=0}^{N} f(x_j) g(x_j),$$

kde za body x_j, $j = 0, \ldots, N$ zvolíme kořeny Čebyševova polynomu T_{N+1} (viz kapitola 6). Dá se ukázat (viz [18]), že

$$\begin{array}{l} (T_j, T_k) = 0, \quad \text{pro } j \neq k,\ 0 \leq j,k \leq N \\ (T_0, T_0) = N+1, \quad (T_j, T_j) = \frac{N+1}{2}, \quad \text{pro } j > 0. \end{array} \tag{10.7}$$

Soustava $(T_0, \ldots, T_N)$ je při daných tabulkových bodech x_j, $j = 0, \ldots, N$, diskrétně ortogonální.

b) Při aproximaci metodou nejmenších čtverců pro ekvidistantní tabulkové body je výhodné užívat, kde je to možné, lichého počtu bodů a prostřední bod brát rovný nule ($N = 2L$). Polynomy, které jsou ortogonální na takové množině bodů se nazývají *Gramovy polynomy* (viz [19]).

Podrobněji: Pro $N = 2L$ definujeme novou proměnnou s předpisem $x = x_L + hs$, tj. $i = L + s$, takže body x_i, $i = 0, \ldots, N$, $x_i = x_0 + ih$ jsou transformovány na celočíselnou množinu $\{-L, \ldots, -1, 0, 1, \ldots, L\}$.

Gramovy polynomy G_j, $j = 0, \ldots, m$, které jsou ortogonální na ekvidistantní množině bodů $\{-L, \ldots, -1, 0, 1, \ldots, L\}$, s krokem $h = 1$, jsou definovány rekurentními vztahy

$$\begin{aligned} &G_0(s) \equiv 1, \quad G_1(s) = s, \\ &G_{j+1}(s) = G_j(s) - \tfrac{j^2((N+1)^2 - j^2)}{4(4j^2-1)} G_{j-1}(s), \quad j \geq 1 \end{aligned} \tag{10.8}$$

Poznámka 4. Z praktických i teoretických důvodů se nedoporučuje používat Gramovy polynomy pro $m > 2\sqrt{N}$.

Příklad 10.2. Pro dané hodnoty nalezněte nejlepší aproximaci pomocí G_0, G_1, G_2.

x_i	0,00	0,25	0,5	0,75	1,00
f_i	1,0000	1,2840	1,6487	2,1170	2,7183

Řešení. Substituce $x = 0,5 + 0,25s$, kde s je nová proměnná, převede počátek souřadnic do bodu $x = 0,5$. Tabulkové body nyní jsou $\{-2, -1, 0, 1, 2\}$ a Gramový polynomy

$$G_0 \equiv 1, \quad G_1(s) = s. \quad G_2(s) = s^2 .$$

Je třeba pouze vypočítat skalární součiny (G_i, G_i), (f, G_i), $i = 0, 1, 2$. Je

$$(G_i, G_i) = \sum_{s=-2}^{2} G_i^2(s) \;\Rightarrow\; (G_0, G_0) = 5,\ (G_1, G_1) = 10,\ (G_2, G_2) = 14 ,$$

dále $(f, G_0) = 8,7680$, $(f, G_1) = 4,2696$, $(f, G_2) = 0,7382$. Normální soustava je tvaru

$$\begin{aligned} 5c_0 &= 8,7680 \\ 10c_1 &= 4,2696 \\ 14c_2 &= 0,7382 \end{aligned}$$

a její řešení je $c_0^* = 1,7536$, $c_1^* = 0,42696$, $c_2^* = 0,05273$. Nejlepší aproximace vyjádřená pomocí Gramových polynomů je

$$P_2(s) = 1,7536 + 0,42696s + 0,05273(s^2 - 2) .$$

Zpětnou substitucí $s = 4(x - 0,5)$ dostaneme vyjádření pomocí proměnné x

$$P_2(x) = 1,0051 + 0,8642x + 0,8437x^2 .$$

Otázkou chyby při aproximaci metodou nejmenších čtverců se zde nebudeme zabývat. Tato problematika je podrobně rozebrána v [19]. Zmíníme se však stručně o významu této metody.

Metodu nejmenších čtverců užíváme obvykle v případech, kdy se jedná o aproximaci funkce, jejíž hodnoty jsou známy pouze empiricky a jsou tedy zatíženy chybami. Aproximace získaná metodou nejmenších čtverců musí mít dvě charakteristické vlastnosti:

1. Musí být dostatečně vysokého stupně, aby aproximující polynom byl dobrou aproximací funkce.

2. Nesmí být příliš vysokého stupně, aby se nepřesnosti v naměřených hodnotách nezachovaly v této aproximaci.

Má-li aproximace získaná metodou nejmenších čtverců tyto dvě vlastnosti, říkáme, že *vyrovnává* naměřené hodnoty v tom smyslu, že *informace* o přesném průběhu funkce se zachová, ale „šum“ se vyrovnáním odstraní. Je-li P_m taková nejlepší aproximace, pak $P_m(x_j)$, $j = 0, \ldots, N$ jsou „vyrovnané“ hodnoty v bodech x_j, $j = 0, \ldots, N$.

V následující tabulce jsou uvedeny vyrovnané hodnoty aproximace P_2 z příkladu 10.2.

i	0	1	2	3	4
x_i	0	0,25	0,50	0,75	1,00
f_i	1,0000	1,2840	1,6487	2,1170	2,7183
$P_2(x_i)$	1,0052	1,2740	1,6482	2,1279	2,7130
$f_i - P_2(x_i)$	-0,0052	0,0100	0,0005	-0,0109	0,0053

Funkce $\rho^2(c_0, c_1, c_2)$ nabývá pro vypočtenou nejlepší aproximaci P_2 své minimální hodnoty a to

$$\rho^2(c_0^*, c_1^*, c_2^*) = \sum_{i=0}^{4} (f_i - P_2(x_i))^2 = 2,76 \cdot 10^{-4}.$$

Cvičení ke kapitole 10

1. Pro hodnoty v příkladě 10.2 nalezněte nejlepší aproximaci tvaru $P_1(x) = c_o + c_1 x$.

   (Řešení: $P_1(x) = l, 70784x + 0, 89968$)

2. Dokažte vztahy (10.7).

3. Metodou nejmenších čtverců najděte polynom stupně druhého, který aproximuje tyto hodnoty:

x_i	-3	-2	-1	0	1	2	3
f_i	4	2	3	0	-1	-2	-5

   a) zvolte nejdřív $\phi_i(x) = x^i$, $i = 0, 1, 2$
   b) užijte Gramovy polynomy G_0, G_1, G_2.

   (Řešení: $P_2(x) = \frac{1}{64}(56 - 117x - 11x^2)$.)

4. (a) Pro hodnoty v příkladě 10.2 užijte tříbodové formule a vypočtěte derivace v bodech 0,25; 0,5; 0,75 (viz kapitola 7). V každém případě položte střed formule do toho bodu, v němž počítáte derivaci.

(b) Užijte aproximace získané metodou nejmenších čtverců v příkladě 10.2 a vypočtěte derivace P_2 v uvedených bodech.

(c) Porovnejte výsledky s hodnotami derivace přesné funkce $f(x) = e^x$.

Literatura

[1] Bachvalov, N. S.: Čislennyje metody. Nauka, Moskva, 1973.

[2] Berezin, I. S., Židkov, N. P.: Metody vyčislenněj I, II. Nauka, Moskva, 1966.

[3] Brandts, J., Křížek M.: Padesát let metody sdružených gradientů. PMFA, 47 (2), 2002, str. 103–113

[4] Burden, R. L., Faires, J. D.: Numerical Analysis. Prindle, Weber and Schmidt, Boston, 1984.

[5] Burlisch, R., Stoer, J.: Introduction to Numerical Analysis. Springer Verlag, New York, Heidelberg, Berlin, 1980.

[6] Datta, B. N.: Numerical Linear Algebra and Applications. ITP, California, 1994.

[7] Hamming, R. W.: Numerical Methods for Scientists and Engineers. McGraw-Hill, New York, 1962.

[8] Isaacson, E., Keller, H. B.: Analysis of Numerical Methods. John Wiley, New York, London, Sydney, 1966.

[9] Jarník, V.: Diferenciální počet (II). Academia, Praha, 1976.

[10] Kobza, J.: Interpolace – vývoj formulace problému a jeho řešení. PMFA, 44 (4), 1999, str. 273–293

[11] Kopal, Z.: Numerical Analysis. Chapman and Hall, London, 1955.

[12] Koukal S., Křížek M., Potůček R.: Fourierovy trigonometrické řady a metoda konečných prvků. Academia, Praha, 2002

[13] Mathews, J. H.: Numerical Methods for Mathematics, Science and Engineering. Prentice-Hall International, Inc., New Jersey, 1992.

[14] Mathews, J. H., Fink, K. D.: Numerical Methods Using MATLAB, Pearson, New Jersey, 2004.

[15] Mikloško, J.: Syntéza a analýza efektívnych numerických algoritmov. Veda, Bratislava, 1979.

[16] Natanson, J. P.: Konstruktivnaja teorija funkcij. Nauka, Moskva, 1949.

[17] Ortega, J. M., Rheinboldt, W. C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, London, 1970.

[18] Přikryl, P.: Numerické metody matematické analýzy. SNTL, Praha, 1988.

[19] Ralston, A.: Základy numerické matematiky. Academia, Praha, 1973.

[20] Smítal, J.: O funkciach a funkcionalnych rovniciach. Alfa, Bratislava, 1984.

[21] Smith, H. V.: Numerical Methods of Integration. Chart.-Bratt Ltd., 1993.

[22] Szegö: Orthogonal Polynomials. AMS, Providence, 1991.

[23] Ševčuk, I. A.: Približenije mnogočlenami i sledy nepreryvnych na otrezke funkcij. Naukova dumka, Kijev, 1992.

[24] Šotová, J.: Cykly v iteračních metodách pro řešení systémů lineárních rovnic. Disertační práce, 1997.

Rejstřík