

Porovnání některých numerických algoritmů

prof. RNDr. Miroslav Vlček, DrSc.

`vlcek@fd.cvut.cz`

13. listopadu 2007

1. Lineárně konvergující algoritmy
2. Aritmeticko-geometrický průměr
3. Kvadraticky konvergující algoritmy
4. Numerické výpočty pro některé speciální funkce

Lineárně konvergující algoritmy

Za domácí úkol jste měli ukázat, jak pomocí řetězového zlomku

$$q_0 + \frac{p_1}{q_1 + \frac{p_2}{q_2 + \frac{p_3}{q_3 + \frac{p_4}{q_4 + \dots}}}} \quad (1)$$

a jeho ekvivalentního zápisu pro čitatele jmenovatele k -tého součtu konečného řetězového zlomku $S_k = \frac{A_k}{B_k}$

$$\begin{bmatrix} A_k & A_{k-1} \\ B_k & B_{k-1} \end{bmatrix} = \begin{bmatrix} q_0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} q_1 & 1 \\ p_1 & 0 \end{bmatrix} \cdots \times \begin{bmatrix} q_k & 1 \\ p_k & 0 \end{bmatrix}$$

obdržíme ze součinu matic algoritmus pro výpočet odmocniny.

Například pro $\sqrt{2}$ platí

$$\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \dots}}}} \quad (2)$$

V tomto případě jsou vektory

$$\overrightarrow{p} = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ \dots],$$

$$q_0 = 1$$

a

$$\overrightarrow{q} = [2 \ 2 \ 2 \ 2 \ 2 \ \dots],$$

takže násobení matic má jednoduchý tvar

$$\begin{bmatrix} A_k & A_{k-1} \\ B_k & B_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}^k$$

Mocnina matice 2×2

Hledejme mocninu matice 2×2 , která má jednotkový determinant $ad - bc = 1$

$$\mathcal{A}^n = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^n$$

Vlastní čísla $\lambda_{1,2}$

$$\det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = \lambda^2 - (a + d)\lambda + ad - bc = 0$$

V dalším budeme značit

$$\Delta = ad - bc = 1$$

$$2x = a + d$$

$$2y = a - d$$

a obdržíme reciprokou kvadratickou rovnici

$$\lambda^2 - (a + d)\lambda + ad - bc = \lambda^2 - 2x\lambda + 1 = 0$$

s řešením

$$\lambda_{1,2} = x \pm \sqrt{x^2 - 1}.$$

Mocnina matice $\mathcal{A}$

$$\mathcal{A}^n = \lambda_1^n P_1 + \lambda_2^n P_2,$$

kde

$$P_1 = \frac{\mathcal{A} - \lambda_2}{\lambda_1 - \lambda_2}$$

a

$$P_2 = \frac{\mathcal{A} - \lambda_1}{\lambda_2 - \lambda_1}$$

jsou tzv. projektory s vlastnostmi

$$P_1 + P_2 = 1,$$

$$P_1 \times P_2 = 0,$$

$$\det P_1 = \det P_2 = 0.$$

Algebraickými úpravami obdržíme

$$P_1 = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} y & b \\ c & -y \end{bmatrix},$$

$$P_2 = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} y & b \\ c & -y \end{bmatrix}.$$

$$\begin{aligned}
\mathcal{A}^n &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{\lambda_1^n + \lambda_2^n}{2} + \begin{bmatrix} y & b \\ c & -y \end{bmatrix} \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} T_n(x) + \begin{bmatrix} y & b \\ c & -y \end{bmatrix} U_{n-1}(x),
\end{aligned}$$

kde $T_n(x)$ a $U_{n-1}(x)$ jsou Čebyševovy polynomy prvního a druhého druhu.

Mocninu matice

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}^k$$

Ize tedy vyjádřit v uzavřeném tvaru

$$\begin{aligned} \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}^k &= (-i)^k \begin{bmatrix} 2i & i \\ i & 0 \end{bmatrix}^k = \\ &= (-i)^k \begin{bmatrix} T_k(i) + iU_{k-1}(i) & iU_{k-1}(i) \\ iU_{k-1}(i) & T_k(i) - iU_{k-1}(i) \end{bmatrix}, \end{aligned}$$

Drobnou algebraickou úpravou zjistíme, že součet konečného řetězového zlomku vyjadřující aproximaci $\sqrt{2}$ má tvar

$$S_{k-1} = \frac{A_{k-1}}{B_{k-1}} = \frac{(-i)^k T_k(i)}{(-i)^{k-1} U_{k-1}(i)}.$$

k	$(-i)^k T_k(i)$	$(-i)^{k-1} U_{k-1}(i)$	S_{k-1}
1	1	1	1.0000000000000000
2	3	2	1.5000000000000000
3	7	5	1.4000000000000000
4	17	12	1.4166666666666667
5	41	29	1.41379310344828
6	99	70	1.41428571428571
7	239	169	1.41420118343195
8	577	408	1.41421568627451
9	1393	985	1.41421319796954
10	3363	2378	1.41421362489487
11	8119	5741	1.41421355164605
12	19601	13860	1.41421356421356
13	47321	33461	1.41421356205732
14	114243	80782	1.41421356242727
15	275807	195025	1.41421356236380
16	665857	470832	1.41421356237469
17	1607521	1136689	1.41421356237282
18	3880899	2744210	1.41421356237314
19	9369319	6625109	1.41421356237309
20	22619537	15994428	1.41421356237310

Přirozená implementace v MATLABu dává následující výsledek.

k	k-tá aproximace $\frac{A_k}{B_k}$	diference $\sqrt{2} - \frac{A_k}{B_k}$
1	1.5000000000000000	-0.08578643762690
2	1.4000000000000000	0.01421356237310
3	1.4166666666666667	-0.00245310429357
4	1.41379310344828	0.00042045892482
5	1.41428571428571	-0.00007215191262
6	1.41420118343195	0.00001237894114
7	1.41421568627451	-0.00000212390141
8	1.41421319796954	0.00000036440355
9	1.41421362489487	-0.00000006252177
	⋮	⋮
19	1.41421356237310	0.0000000000000000

Pro určení odmocniny $\sqrt{2}$ s přesností na 16 platných cifer potřebujeme 19 iteračních kroků.

Jak se změní uvedený algoritmus pro výpočet $\sqrt{3}$?

pro $\sqrt{3}$ platí

$$\sqrt{3} = 1 + \frac{2}{2 + \frac{2}{2 + \frac{2}{2 + \frac{2}{2 + \dots}}}} \quad (3)$$

V tomto případě jsou vektory

$$\overrightarrow{p} \equiv \overrightarrow{q} = [2\ 2\ 2\ 2\ 2\ 2\ \dots],$$

$$q_0 = 1$$

takže násobení matic má jednoduchý tvar

$$\begin{bmatrix} A_k & A_{k-1} \\ B_k & B_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 2 & 0 \end{bmatrix}^k$$

k	k-tá aproximace $\frac{A_k}{B_k}$	diference $\sqrt{3} - \frac{A_k}{B_k}$
1	2.0000000000000000	-0.26794919243112
2	1.6666666666666667	0.06538414090221
3	1.7500000000000000	-0.01794919243112
4	1.7272727272727273	0.00477808029615
5	1.7333333333333333	-0.00128252576446
6	1.73170731707317	0.00034349049571
7	1.73214285714286	-0.00009204957398
8	1.73202614379085	0.00002466377803
9	1.73205741626794	-0.00000660869907
10	1.73204903677758	0.00000177079129
11	1.73205128205128	-0.00000047448240
12	1.73205068043172	0.00000012713716
	:	:
25	1.73205080756888	0.0000000000000000

Pro určení odmocniny $\sqrt{3}$ s přesností na 16 platných cifer potřebujeme 25 iteračních kroků.

Uvedené algoritmy konvergují lineárně.

Kvadraticky konvergující algoritmy

Tento algoritmus vychází z obecného aritmeticko-geometrického průměru. Odmocninu z čísla x nalezneme z rekurentně počítaného aritmetického průměru $\frac{1}{2} \left[y(n-1) + \frac{x}{y(n-1)} \right]$ odhadů odmocniny $y(n-1)$ a podílu $\frac{x}{y(n-1)}$ podle formule

$$y(n) = \frac{1}{2} \left[y(n-1) + \frac{x}{y(n-1)} \right]. \quad (4)$$

Například odmocnina z čísla 10 je s přesností na 14 desetinných míst rovna

$$\sqrt{10} = 3.16227766016838.$$

Pro $x = 10$ dostáváme postupně

$$\begin{aligned} y(1) &= 3 \\ y(2) &= 3.165 \\ y(3) &= 3.162278 \\ y(4) &= 3.1622776601 \\ y(5) &= 3.16227766016835 \\ &\vdots \end{aligned}$$

Aritmeticko-geometrický průměr

Aritmeticko-geometrický průměr dvou kladných čísel $a_0 > b_0$ je velmi prostý a rychle konvergentní algoritmus. Pro aritmetický a geometrický průměr $a_1 = \frac{a_0 + b_0}{2}$, $b_1 = \sqrt{a_0 b_0}$ dvou libovolných kladných čísel a_0 a b_0 platí

$$a_1 = \frac{a_0 + b_0}{2} > \sqrt{a_0 b_0} = b_1. \quad (5)$$

Sestavíme-li nyní posloupnosti $\{a_m\}$, $\{b_m\}$ tak, že

$$a_{m+1} = \frac{a_m + b_m}{2}, b_{m+1} = \sqrt{a_m b_m}, \quad (6)$$

můžeme jednoduchými úpravami ukázat, že tyto dvě posloupnosti konvergují ke společné limitě $M(a_0, b_0)$, neboť platí

$$b_m < b_{m+1} < M(a_0, b_0) < a_{m+1} < a_m.$$

Společnou limitu $M(a_0, b_0)$ nazýváme aritmeticko-geometrickým průměrem čísel a_0 a b_0 . Této hodnoty dosáhneme v každém kroku m s přesností, jejíž mírou je posloupnost $\{c_m\}$

$$c_{m+1} = \frac{a_m - b_m}{2}. \quad (7)$$

V praktických případech počet kroků nepřesáhne jednu desítku.

Výpočet úplných eliptických integrálů

$$\mathbf{K}(k), \mathbf{K}(k'),$$

neúplných eliptických integrálů

$$F(\phi|k)$$

a Jacobiho eliptických funkcí

$$\operatorname{sn}(u|k), \operatorname{cn}(u|k) \text{ a } \operatorname{dn}(u|k)$$

je založen na aritmeticko-geometrického průměru.

Výpočet úplného eliptického integrálu

$$\mathbf{K}(k) = \int_0^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}}$$

Položíme-li

$$a_0 = 1, b_0 = \sqrt{1 - k^2} = k', c_0 = k, \quad (8)$$

potom úplný eliptický integrál modulu k je roven

$$\mathbf{K}(k) = \frac{\pi}{2\mathbf{M}(1, k')} \quad (9)$$

Záměnou

$$a_0 = 1, b_0 = k, c_0 = \sqrt{1 - k^2} = k', \quad (10)$$

dostáváme eliptický integrál doplňkového modulu k'

$$\mathbf{K}(k') = \frac{\pi}{2\mathbf{M}(1, k)} \quad (11)$$

Pro

$k = 0.86602540378444$ a $k' = 0.5000000000000000$

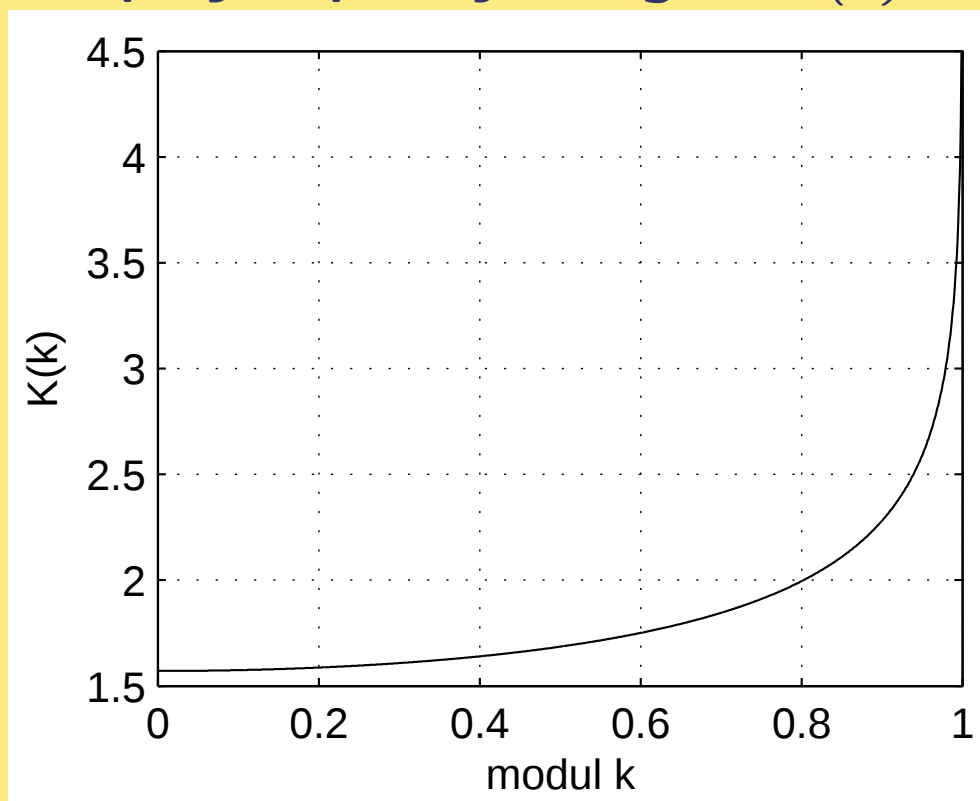
je

m	a_m	b_m
0	1.0000000000000000	0.5000000000000000
1	0.7500000000000000	0.70710678118655
2	0.72855339059327	0.72823765756099
3	0.72839552407713	0.72839550696978
4	0.72839551552345	0.72839551552345

a platí

$$\mathbf{K}(k) = \frac{\pi}{a_4} = 2.15651564749964$$

Úplný eliptický integrál $K(k)$



Aplikace - perioda kyvadla popsaného rovnicí

$$\dot{\vartheta}^2 - 2\omega^2 \cos \vartheta = konst.$$

není $T = \frac{2\pi}{\omega}$, ale $T = \frac{4K(k)}{\omega}$, kde modul eliptického integrálu

$$k = \sin \frac{\vartheta_0}{2}$$

je určen maximální výchylkou kyvadla ϑ_0 .

Výpočet neúplného eliptického integrálu

$$F(\phi|k) = \int_0^\phi \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}$$

1. K zadanému modulu k vypočítáme se zvolenou přesností ϵ posloupnosti $\{a_m\}$, $\{b_m\}$. Algoritmus ukončíme pro jistou hodnotu N , pro kterou již platí

$$c_N < \epsilon \quad (12)$$

2. K zadanému argumentu ϕ vypočítáme vzeštně posloupnost $\{\phi_m\}$ podle rekurentního vztahu

$$\begin{aligned} \phi_0 &= \phi, \\ \phi_{m+1} &= \phi_m + \arctan\left(\frac{b_m}{a_m} \tan \phi_m\right). \end{aligned}$$

Přitom je třeba zaručit, aby $\{\phi_m\}$ byla rostoucí posloupnost ($\phi_{m+1} \sim 2\phi_m$). Potom

$$F(\phi|k) = \frac{\phi_N}{2^N a_N} . \quad (13)$$

Pro

$$\begin{aligned} k &= 0.86602540378444 \\ \phi = \pi/3 &= 1.04719755119660 \end{aligned}$$

je

m	a_m	ϕ_m
0	1.0000000000000000	1.04719755119660
1	0.7500000000000000	1.76092193014136
2	0.72855339059327	3.53307617772659
3	0.72839552407713	7.06599949939667
4	0.72839551552345	14.13199898705032

a platí

$$F(\pi/3|k) = \frac{\phi_4}{2^4 a_4} = 1.21259661525498$$

Výpočet Jacobiho eliptických funkcí

Jacobiho eliptické funkce jsou vlastně inverzními funkcemi eliptických integrálů a proto algoritmus jejich výpočtu je obdobný předcházejícímu.

1. K zadanému modulu k vypočítáme se zvolenou přesností ϵ posloupnosti $\{a_m\}$, $\{b_m\}$, $\{c_m\}$.

2. Nalezneme reálnou čtvrtperiodu

$$K(k) = \frac{\pi}{2a_N} \quad (14)$$

a k zadanému argumentu u určíme

$$\phi_N = 2^N a_N u. \quad (15)$$

3. Z rekurentní formule

$$\phi_{m-1} = \frac{1}{2}[\phi_m + \arcsin(\frac{c_m}{a_m} \sin \phi_m)], \quad (16)$$

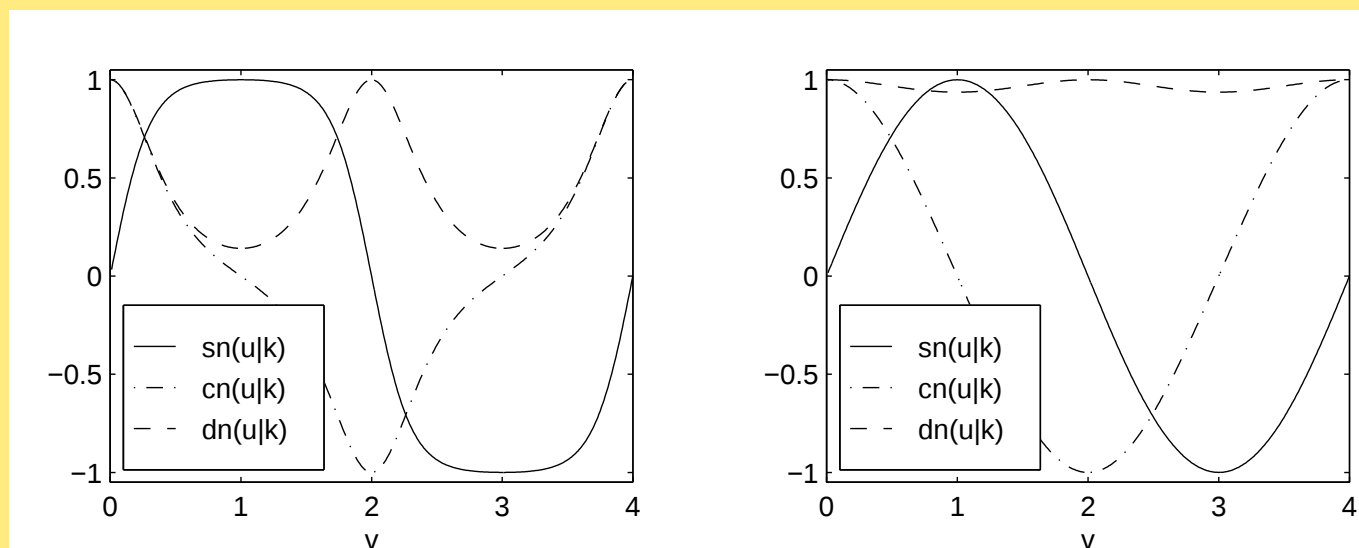
nalezneme sestupně posloupnost $\{\phi_m\}$. Eliptické funkce $\operatorname{sn}(u|k)$, $\operatorname{cn}(u|k)$ a $\operatorname{dn}(u|k)$ jsou poto určeny vztahy

$$\begin{aligned}\operatorname{sn}(u|k) &= \sin \phi_0, \\ \operatorname{cn}(u|k) &= \cos \phi_0, \\ \operatorname{dn}(u|k) &= \frac{\cos \phi_0}{\cos(\phi_1 - \phi_0)}.\end{aligned}$$

Pro hodnotu argumentu rovnou lichým násobkům $\mathbf{K}(k)$ je $\cos \phi_0 = 0$ a funkci $\operatorname{dn}(u|k)$ počítáme ze vztahu

$$\operatorname{dn}(\mathbf{K}(k)|k) = k'. \quad (17)$$

Jacobiho eliptické funkce



Uvedené algoritmy konvergují kvadraticky.

Rekurentní algoritmy pro Besselovy funkce

Besselovy funkce 1. druhu $J_n(x)$ lze vyjádřit ve tvaru nekonečné řady

$$J_n(x) = \left(\frac{x}{2}\right)^n \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(n+k+1)} \left(-\frac{x^2}{4}\right)^k \quad (18)$$

Besselovy funkce splňují rekurentní formuli

$$J_{n+1}(x) + J_{n-1}(x) = \frac{2n}{x} J_n(x) \quad (19)$$

Asympticky se $J_n(x)$ jako funkce řádu n blíží k nule, tj., že pro zvolenou hodnotu x nalezneme n vždy takové, že platí

$$J_n(x) \sim \frac{1}{\sqrt{2\pi n}} \left(\frac{x e}{2n}\right)^n < \epsilon \quad (20)$$

Tato asymptotická vlastnost říká, na kolik platných cifer je pro dané x a n splněna rovnost

$$J_{n+1}(x) = 0.$$

Numerický výpočet $J_n(x)$

1. K zadané hodnotě argumentu x a přesnost ϵ , například $\epsilon = 10^{-8}$ nalezneme podle (20) hodnotu n a položíme

$$\begin{aligned} F_{n+1} &= 0 \\ F_n &= 1 \end{aligned}$$

2. Vypočteme sestupně posloupnost $\{F_k\}$ z rekurentní formule

$$F_{k-1} = \frac{2n}{x} F_k - F_{k+1} \quad (21)$$

3. Prvky této posloupnosti jsou již úměrné Besselovým funkcím argumentu x , tj.

$$J_k(x) = \frac{1}{C} F_k \quad (22)$$

4. Normovací konstantu určíme ze vztahu

$$C = F_0 + 2 \sum_{k=1}^{[n/2]} F_{2k}, \quad (23)$$

který je obdobou nekonečné sumy

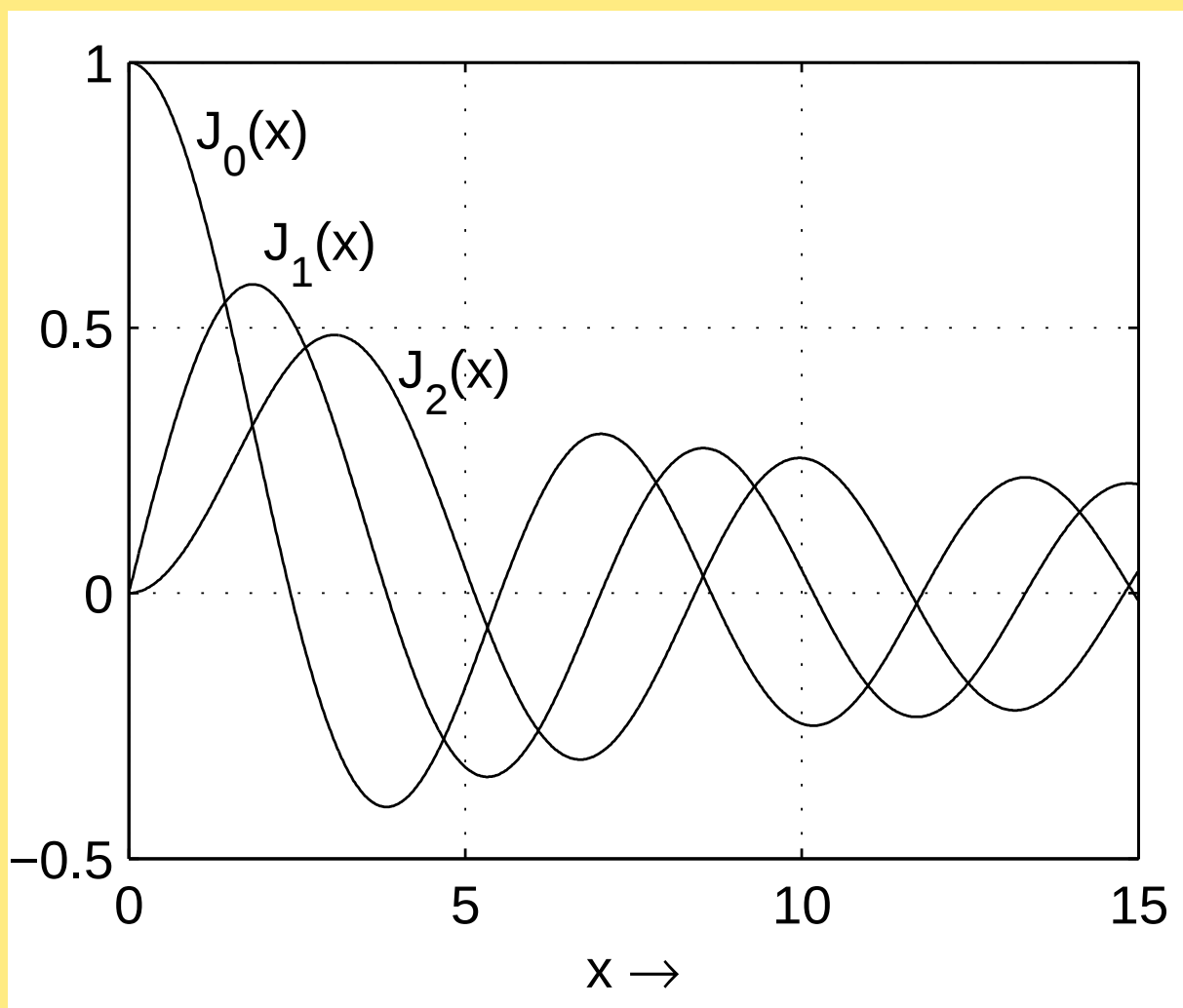
$$1 = J_0(x) + 2 \sum_{k=1}^{\infty} J_{2k}(x), \quad (24)$$

Tím je algoritmus ukončen. Požadovaná přesnost je splněna, srovnání s tabelovanými hodnotami vychází velmi dobře.

Zpětná a dopředná rekurence pro $J_n(1)$

n	F_n	$J_n(x)$ zpětná	$J_n(x)$ dopředná
0	63994242408	0.76519768655	0.76519768655
1	36801867479	0.44005058574	0.44005058574
2	9609492550	0.11490348493	0.11490348493
3	1636102721	0.01956335398	0.01956335398
4	207123776	0.00247663896	0.00247663896
5	20887487	0.00024975773	0.00024975773
6	1751094	0.00002093833	0.00002093833
7	125641	0.00000150232	0.00000150232
8	7880	0.00000009422	0.00000009420
9	439	0.00000000524	0.000000000499
10	22	0.00000000026	-0.000000000423
11	1	0.00000000001	-0.000000008963

První tři Besselovy funkce $J_k(x)$.



Děkuji za pozornost

