Az integrálszámítás középértéktétele szerint

$$R_1(f)=\frac{f''(\eta)}{2}\int_a^b (x-c)^2dx=\frac{f''(\eta)}{2}\left[\frac{(x-c)^3}{3}\right]_a^b=\frac{(b-a)^3}{24}f''(\eta). \qquad \blacksquare$$

A gyakorlatban ezt a formulát nem az egész $[a,b]$ intervallumra alkalmazzuk, hanem azt m részre osztjuk, és az egyes részintervallumokban az érintőformulával integrálunk. Például, ha $m=3$: $h=(b-a)/3$ és a három részintervallumra alkalmazzuk a (18.7) szabályt.

A részintervallumon nyert eredmények felösszegzésével jutunk az *érintőszabályhoz:*

$$\int_a^b f=\frac{b-a}{m}\sum_{i=1}^{m} f(a-h/2+ih)+\frac{(b-a)^3}{24m^2}f''(\eta), \tag{18.9}$$

ahol most $f''(\eta)=\frac{1}{m}(f''(\eta_1)+f''(\eta_2)+\ldots+f''(\eta_m))$, mert a Darboux-tulajdonság miatt f'' ezt az átlagértéket is felveszi valahol a teljes intervallumban.

2. A *trapézformula*. Elsőfokú polinom interpolációból nyert zárt formula: $n=1$, $B_0^z=B_1^z=1/2$:

$$I_1(f)=\frac{b-a}{2}(f(a)+f(b)). \tag{18.10}$$

18.2.2 Tétel, trapéz formula hibája

Legyen $f\in C^2[a,b]$, *ekkor*

$$\int_a^b f=\frac{b-a}{2}(f(a)+f(b))-\frac{(b-a)^3}{12}f''(\eta), \quad \eta\in[a,b]. \tag{18.11}$$

<u>Bizonyítás.</u> Az interpoláció hibatagjának integrálja az integrálszámítás középérték tételének felhasználásával:

$$R_1(f)=\int_a^b \frac{f''(\xi_x)}{2}(x-a)(x-b)dx=-\int_a^b \frac{f''(\xi_x)}{2}\underbrace{(x-a)(b-x)}_{\geq 0}dx=-\frac{f''(\eta)}{12}(b-a)^3. \qquad \blacksquare$$

A teljes intervallumot m részre osztva, a részintervallumok eredményét felösszegezve nyerjük a *trapéz szabályt*:

$$\int_a^b f=\frac{b-a}{2m}[f(x_0)+2f(x_1)+\ldots+2f(x_{m-1})+f(x_m)]-\frac{(b-a)^3}{12m^2}f''(\eta). \tag{18.12}$$

18.2.3 Definició

Egy kvadratúra formulát akkor mondunk *k-adrendűnek,* ha k-adfokú az a legkisebb fokszámú polinom, amelyre a formula már nem pontos.

Eszerint az érintő- és trapézformula ***másodrendű.***

3. A *Simpson formula*: másodfokú polinom interpolációból nyert zárt formula, $n=2$, $B_0^z=1/6$, $B_1^z=4/6$, $B_2^z=1/6$ és

$$I_2(f)=\frac{b-a}{6}(f(a)+4f(\frac{a+b}{2})+f(b)).$$

Az interpoláció maradéktagjában $\omega_2(x)=(x-a)(x-\frac{a+b}{2})(x-b)$ szerepel, ennek az integrálja $[a,b]$-ben zérus. Ezt legegyszerűbben úgy tudjuk belátni, hogy $[a,b]$-t a $[-1,1]$ intervallumba transzformáljuk. Ekkor $\omega_2(x)$ páratlan függvény, amelynek az integrálja zérus. Emiatt a hibatételt az Hermite-interpolációból származtatjuk,

$$f(x)=H_3(x)+\frac{f^{(4)}(\xi_x)}{4!}(x-a)(x-\frac{a+b}{2})^2(x-b), \tag{18.13}$$

ahol az $(a+b)/2$ középpontban az első deriváltat is interpoláljuk. Az általánosított osztott differenciák táblázatára gondolva tudjuk, hogy az interpoláló polinom a következő alakú: $H_3(x)=L_2(x)+C(x-a)(x-\frac{a+b}{2})(x-b)$. A második tag együtthatójának értéke nem fontos, mert az integrálja az előbbiek alapján zérus, s ezzel $\int_a^b H_3=\int_a^b L_2$.

18.2.4 Tétel, Simpson-formula hibája

Legyen $f\in C^4[a,b]$. *Ekkor létezik* $\eta\in[a,b]$, *amelyre*

$$\begin{aligned}\int_a^b f&=\frac{b-a}{6}\left(f(a)+4f(\frac{a+b}{2})+f(b)\right)-\frac{f^{(4)}(\eta)}{90}\left(\frac{b-a}{2}\right)^5=\\&=I_2(f)-\frac{f^{(4)}(\eta)}{90}h^5,\end{aligned} \tag{18.14}$$

ahol $h=(b-a)/2$.

<u>Bizonyítás.</u> Kiindulunk az Hermite-interpoláció (18.13) alakjából, ahonnan integrálással kapjuk:

$$I(f)-I_2(f)=\int_a^b \frac{f^{(4)}(\xi_x)}{4!}(x-a)(x-\frac{a+b}{2})^2(x-b)dx.$$

Ahhoz, hogy az integrálszámítás középértéktételét alkalmazhassuk, az $f^{(4)}$ mellett álló tényező nem lehet negatív. Ezt úgy biztosíthatjuk, hogy $(x-b)$ helyett $(b-x)$-et írunk, s így

$$I(f)-I_2(f)=-\frac{f^{(4)}(\eta)}{4!}\int_a^b (x-a)(x-\frac{a+b}{2})^2(b-x)dx=-\frac{f^{(4)}(\eta)}{90}\left(\frac{b-a}{2}\right)^5. \qquad \blacksquare$$

Simpson-szabály. A teljes $b-a$ intervallumot páros számú m részintervallumra osztva és a Simpson-formulát a szomszédos intervallumpárokra alkalmazva kapjuk a Simpson-szabályt, mint összetett formulát. Ekkor három pontonként fogjuk össze a formulákat és az összetett formula:

$$\begin{aligned}\int_a^b f&=\sum_{k=1,3,\ldots}\left(\frac{2h}{6}(f(x_{k-1})+4f(x_k)+f(x_{k+1}))-\frac{f^{(4)}(\eta_k)}{90}h^5\right)=\\&=\frac{h}{3}\left(f(x_0)+2\sum_{\substack{k\ \text{páros}\\ \text{belső pont}}}f(x_k)+4\sum_{k\ \text{pálan}}f(x_k)+f(x_m)\right)-\frac{h^5}{90}\sum_{k\ \text{pálan}}f^{(4)}(\eta_k).\end{aligned} \tag{18.15}$$

A hibatag még tovább írható:

18. Numerikus integrálás (kvadratúra) I.

Az integrálok kiszámításakor nem mindig ismert a primitív függvény, vagy ha igen, némely esetben nagyon bonyolult, nehezen számítható. Ilyenkor a numerikus módszerek a kívánt pontosságú eredmény előállítására egyszerűbb alternatívát kínálnak. A továbbiakban az interpolációból nyerhető kvadratúra-formulákkal fogunk foglalkozni.

Láttuk, a függvény az $[a,b]$ intervallumban a következő módon állítható elő:

$$f = L_n + r_n, \tag{18.1}$$

ahol L_n a Lagrange-interpolációs polinom és r_n a hibatag. (Feltesszük, az alappontok nagyság szerint rendezettek: $x_{i-1} < x_i$ és $x_0 = a$, $x_n = b$.) A kvadratúra-formulák származ-tatási elve:

$$\int_a^b f = \int_a^b L_n + \int_a^b r_n = \sum_{i=0}^{n} a_i f(x_i) + R_n, \tag{18.2}$$

ahol az

$$a_i = \int_a^b l_i(x)dx \tag{18.3}$$

súlyok a Lagrange alappolinomok integrálásából adódnak.

Következmény. Az így nyert formulák legfeljebb n-edrendű polinomig pontosak.

Ekvidisztáns alappontok esetén nyerjük a Newton-Cotes formulákat.

18.1. Zárt és nyilt Newton-Cotes kvadratúra formulák

18.1.1 Definició

Az alappontok halmaza legyen $\Omega_n = \{x_0,\dots,x_n\}$. *Zárt* a kvadratúra-formula, ha $a,b \in \Omega_n$, $h = (b-a)/n$, $x_k = a + k\cdot h$, $k = 0,1,\dots,n$. *Nyilt* a formula, ha $a,b \notin \Omega_n$, $h = (b-a)/(n+2)$, $x_k = a + (k+1)\cdot h$, $k = 0,1,\dots,n$, $x_{-1} = a$, $x_{n+1} = b$.

A továbbiakban rátérünk a zárt Newton-Cotes formulák együtthatóinak előállítására.

$$a_k = \int_a^b l_k(x)dx = \int_a^b \frac{\omega_n(x)}{(x-x_k)\omega_n'(x_k)}dx.$$

Vegyük észre: $x_k - x_j = (k-j)h$ és vezessünk be új változót: $t = (x-a)/h$, ahonnan $x = a + th$ és $x - x_j = (t-j)h$, s ezzel

$$\begin{aligned} a_k &= \int_0^n \frac{t(t-1)\overbrace{\dots\dots\dots}^{(t-k)\text{ hiányzik}}(t-n)h^n}{k(k-1)\dots1(-1)(-2)\dots(-n+k)h^n}hdt = \\ &= (b-a)\frac{(-1)^{n-k}}{nk!(n-k)!}\int_0^n \frac{t(t-1)\dots(t-n)}{t-k}dt = (b-a)B_{k,n}^{\text{zárt}}, \end{aligned} \tag{18.4}$$

ahol a $B_{k,n}^{\text{zárt}}$ együtthatók az intervallumtól függetlenül egyszer s mindenkorra kiszámíthatók. Hasonló módon nyerhetjük a *nyilt Newton-Cotes* formulák együtthatóit:

19. Numerikus integrálás, Gauss-kvadratúrák II.

Az eddigi, interpolációból származtatott kvadratúra-formulák legalább annyiad fokú polinomra pontosak, ahányad fokú polinomból származtattuk őket. A Gauss-kvadratúrák abból az észrevételből származnak, hogy az alappontok speciális megválasztásával a kavadratúra-formula rendje növelhető. Ismét szükségünk lesz az ortogonális polinomokra.

19.1. Tétel, ortogonális polinom gyökei

Legyen $\{p_k(x)\}$ egy ortogonális polinom rendszer. Ekkor bármely n-re $p_{n+1}(x)$ gyökei valósak, egyszeresek és az $[a,b]$ intervallumban vannak, ahol $[a,b]$ a skalárszorzat integrálási tartománya.

Bizonyítás. Legyenek $x_0,x_1,\dots,x_k$ $p_{n+1}(x)$ páratlan multiplicitású gyökei $[a,b]$-ben, azaz ott $p_{n+1}(x)$ előjelet vált. Ha $k = n$, akkor a tétel állítása rendben van. Ha nem, akkor indirekt úton feltesszük: $k < n$ és megmutatjuk, hogy az állítás ellentmondásra vezet. Ehhez tekintsük a $q(x) = (x-x_0)(x-x_1)\dots(x-x_k)$ $k+1$-edfokú polinomot. Mivel $k+1 < n+1$, az ortogonalitás miatt $(p_{n+1},q) = 0$. De ezzel ellentmondásra jutunk, mert $p_{n+1}(x)q(x)$ nem vált előjelet $[a,b]$-ben, mivel p_{n+1} minden előjelválátását $q(x)$ megszünteti és így $\int p_{n+1}q\alpha \neq 0$ volna. Vegyük észre, a gondolatmenet akkor is jó, ha egyetlen páratlan multiplicitású gyök sincs, mert ekkor $q(x)$ 0-adfokú.

■

Az $n+1$-pontos Gauss-kvadratúrát úgy kapjuk, hogy a $p_{n+1}(x)$ ortogonális polinom gyök-helyein készítjük az interpolációból származtatott kvadratúra-formulát. A séma a következő:

$$\int_a^b f\alpha = \int_a^b L_n\alpha + \int_a^b r_n\alpha = \sum_{i=0}^{n} a_i f(x_i) + R_n, \quad a_i = \int_a^b l_i(x)\alpha(x)dx. \tag{19.1}$$

19.2. Tétel, Gauss-kvadratúra pontossága

Legyenek a $p_{n+1}(x)$ ortogonális polinom gyökei $x_0,x_1,\dots,x_n$, $a_i = \int l_i\alpha$, ahol l_i az i-edik Lagrange alappolinom a fenti alappontokon. Ekkor a Gauss-kvadratúra

$$G_n(f) = \sum_{i=0}^{n} a_i f(x_i)$$

pontos minden legfeljebb $2n+1$-edfokú polinomra: $f \in \mathcal{P}_{2n+1} \to \int f\alpha = G_n(f)$.

Bizonyítás. Az interpolációból való származatatás miatt $G_n(f)$ biztosan pontos a legfeljebb n-edfokú polinomokra. Tegyük fel, $f \in \mathcal{P}_{2n+1}$, $f = p_{n+1}\cdot q + r$, $q,r \in \mathcal{P}_n$, így

$$\begin{aligned} G_n(f) &= \sum_{i=0}^{n} a_i f(x_i) = \sum_{i=0}^{n} a_i[\underbrace{p_{n+1}(x_i)}_{=0\text{ minden }i\text{-re}}\cdot q(x_i) + r(x_i)] = \\ &= \sum_{i=0}^{n} a_i r(x_i) = G_n(r) = \int r\alpha = \qquad (\text{mert } n\text{-edfokig pontos}) \\ &= \int (p_{n+1}\cdot q + r)\alpha = \qquad (\text{mert } q \in \mathcal{P}_n, \text{ így ortogonális } p_{n+1}\text{-re}) \\ &= \int f\alpha. \end{aligned}$$

A (17.6) alak összefüggésbe hozható az un. *Frobenius-féle kisérő mátrix*-szal, amellyel már találkoztunk a 7.3 szakaszban:

$$F=\begin{bmatrix} 0 & 0 & \dots & \dots & 0 & -a_0/a_n \\ 1 & 0 & \dots & \dots & 0 & -a_1/a_n \\ & 1 & \ddots & & \vdots & \vdots \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & \ddots & 0 & -a_{n-2}/a_n \\ & & & & 1 & -a_{n-1}/a_n \end{bmatrix}. \tag{17.7}$$

Az utolsó oszlopa mentén kifejtve könnyen igazolható, hogy $\det(\lambda I-F)=\frac{1}{a_n}\sum_{i=0}^{n}a_i\lambda^i$. Ennek a mátrixnak ismerete két szempontból is hasznos. Egyrészt mutatja, hogy a polinom x_k gyökei lineáris algebrai módszerekkel kereshetők, amelyek a legstabilabbnak tekinthető módszerek közé tartoznak. Másrészt rögtön lehetőségünk van egy olyan körlemez megadására a komplex síkon, amelyben a polinom összes gyöke benne van:

$$\|F\|_\infty=\max_{0\le i<n}\left(1-\delta_{i0}+|a_i/a_n|\right)=R\ge|x_k|,$$

ahol δ_{ij} a Kronecker delta. R nagyobb vagy egyenlő F spektrál sugaránál, ami most a polinom gyökök abszolút értékeinek maximuma.

Megadhatunk egy másik kisebb körlemezt is, amelyen kívül van az összes gyök. Vezessük be az $x=1/y$ transzformációt és írjuk át a polinomot y szerint. Eredményül egy olyan polinomot kapunk, ahol az együtthatók fordított sorrendben vannak és ennek a polinomnak a gyökei az eredeti gyökök reciprokai. Az új polinomhoz tartozó Frobenius-féle mátrix sornormáját véve kapjuk: $1/|x_k|\le\max_{0<i\le n}\left(1-\delta_{0i}+|a_i/a_0|\right)=1/r$, ahol természetesen feltételeztük, hogy $a_0\ne 0$. A két eredményt egybevetve látjuk, hogy a polinom gyökei az

$$r\le|x_k|\le R,\quad k=1,2,\dots,n \tag{17.8}$$

körgyűrű tartományba esenek.

A (17.6)-tal adott polinomoknál előnyösen alkalmazható a Newton-módszer, mert a polinom értéke és a deriváltja egy lépésben egyszerűen számolható. Ha például a ξ helyen szeretnénk ezeket kiszámolni, nem kell mást tennünk, mint a polinomot maradékos osztással elosztani $(x-\xi)^2$-tel:

$$p(x)=q(x)(x-\xi)^2+\alpha x+\beta. \tag{17.9}$$

Könnyen meggyőződhetünk róla, hogy a helyettesítési érték $\alpha\xi+\beta$, a derivált pedig α lesz a ξ helyen.

A többszörös gyökök kiszűrésére alkalmazhatjuk az Euklidészi algoritmust. Ekkor a két induló polinom $p_0(x)=p(x)$, $p_1(x)=-p'(x)$, az $i+1$-edik polinomot pedig úgy készítjük, hogy $p_{i-1}(x)$-et osztjuk $p_i(x)$-szel és a maradékot képezzük:

$$p_{i-1}(x)=q_i(x)p_i(x)-c_ip_{i+1}(x),\quad i=1,2,\dots \tag{17.10}$$

A sorozatban a polinomok fokszáma csökkenő, $c_i>0$, egyébként tetszőleges. Az algoritmus $m\le n$ lépés után befejeződik:

$$p_{m-1}(x)=q_m(x)p_m(x),\quad p_m(x)\ne 0.$$

Az utolsó polinom a két kezdő polinom legnagyobb közös osztója. Mivel a derivált polinom az 1-nél nagyobb multiplicitású gyököket tartalmazza, így ezek a gyökök megjelennek $p_m(x)$-ben.

19.3. Példák

1. Három-pontos Gauss-Csebisev kvadratúrával közelítjük a $\int_{-1}^{1}\left(1-x^2\right)^{-1/2}e^{-x}dx$ integrált. Becsüljük meg a hibát!

Megoldás. A három-pontos kvadratúránál $n=2$, ezt alkalmazzuk a (19.2) formulánál: $M_6=e$ és mivel 1-főegyütthatós polinomoknak kell szerepelni, emiatt $p_3(x)=T_3(x)/4$. Így a hiba kisebb, mint $e\cdot(T_3,T_3)/(16\cdot 6!)=e\pi/(32\cdot 720)$, mert $(T_3,T_3)=\pi/2$, (lásd a 7.4 feladatot).

2. Készítsük el a két-pontos Gauss-Hermite kvadratúra súlyait! Ellenőrízzük, hogy az így kapott kvadratúra legfeljebb harmadfokú polinomokra pontos!

Megoldás. A két-pontos kvadratúránál a másodfokú Hermite-polinom gyökei $-x_0=x_1=2^{-1/2}$. Így $a_0=\int_{-\infty}^{\infty}\frac{(x-x_1)}{(x_0-x_1)}\exp(-x^2)dx=\frac{x_1\mu_0}{2x_1}=\mu_0/2$, mert az elsőfokú tag integrálja zérus, lévén az integrandus páratlan függvény. Hasonlóan adódik, hogy $a_1=a_0$. A kapott kvadaratúra pontos az 1 függvényre, mert az eredmény μ_0. Az x és x^3 függvényre is pontos, mert a két tag a gyökhelyeken a páratlanság miatt kiejti egymást. Így már csak azt kell igazolunk, hogy a pontosság a másodfokú x^2 polinomra is teljesül. Ennek az integrálja a (9.8) formula alapján nem más mint $(p_1,p_1)=\beta_1(p_0,p_0)$, ami az előző oldalon látható táblázat szerint egyenlő $\mu_0/2$-vel. A Gauss-Hermite kvadratúra eredménye pedig $\frac{\mu_0}{2}(\frac{1}{2}+\frac{1}{2})$, tehát igaz az egyezés.

3. Határozzuk meg a következő integrál értékét:

$$\int_{-1}^{1}\frac{(2x^2+x)dx}{\sqrt{1-x^2}}.$$

Megoldás. A számlálóban szereplő polinomot előállítjuk az első három Csebisev polinom lineáris kombinációjaként: $2x^2+x=c_0T_0+c_1T_1+c_2T_2$. Ezt felhasználva az integrálunk így is írható: $(T_0,c_0T_0+c_1T_1+c_2T_2)=c_0(T_0,T_0)=c_0\pi$. Az első három Csebisev polinom együtthatóival a következő lineáris egyenletrendszert írhatjuk fel (9.5) alapján:

$$\begin{bmatrix} 1 & 0 & -1 \\ & 1 & 0 \\ & & 2 \end{bmatrix}\begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix},$$

amiből $c_0=1$, tehát az integrál értéke π.

9. Ellenőrízzük, hogy a Newton-módszer többszörös gyök esetén csak elsőrendben konvergál.
10. Igazoljuk, hogy r-szeres multiplicitású gyöknél a kvadratikus konvergencia megmarad, ha a Newton-módszer formuláját a következőre módosítjuk: $x_{n+1} = x_n - rf(x_n)/f'(x_n)$.
11. Adott ε pontosság elérése érdekében dolgozzuk ki annak feltételét, hogy mikor állítsuk le a Newton-módszert.
12. Mi történik a szelőmódszernél, ha a 16.4.3 Tétel feltételeitől csak annyi a különbség, hogy $\varepsilon_0 > 0$, de $\varepsilon_1 < 0$?
13. Mikor állítsuk le a szelőmódszert, hogy a megoldás előírt pontosságú legyen?

17. Nemlineáris egyenletek megoldása II.

Néhány speciális esettel folytatjuk.

17.1. Az intervallumfelezés módszere

Tegyük fel, az $[a,b]$ intervallum tartalmaz 1 db gyököt: $f(a)f(b)<0$ és a függvény folytonos $[a,b]$-ben. Az intervallumfelezés módszere szerint ekkor megfelezzük az intervallumot és a két intervallum közül megtartjuk azt, ahol az előjelváltás megmarad. Így az algoritmus:

1. $\exists f \in C[a,b]$, $f(a)f(b)<0$ és adott $\in$ előírt pontosság.
2. Indulás: $[a_0,b_0]=[a,b]$, $x_1=(a+b)/2$.
3. $[a_n,b_n]=\begin{cases}[a_{n-1},x_n], & \text{ha } f(a_{n-1})f(x_n)<0,\\ [x_n,b_{n-1}], & \text{egyébként,}\end{cases}$
   $x_{n+1}=(a_n+b_n)/2.$
4. Megállás: ha $f(x_n)=0$, vagy $|b_n-a_n|<\in$.

Ez nem túl gyors, de biztos módszer. Az előjelváltásból nem mindig következik a gyök léte. Gondoljunk az $1/x$ függvényre, amikor az algoritmust -1 és 2 között indítjuk.

17.1.1 Tétel

Az intervallumfelezéssel kapott x_n, $n=1,2,\ldots$ sorozat elsőrendben konvergens és

$$|\varepsilon_n| \le \frac{b-a}{2^n}, \quad n=0,1,\ldots \tag{17.1}$$

<u>Bizonyítás.</u> A konvergencia abból következik, hogy mindig a gyököt tartalmazó intervallumot tartjuk meg. A hibára minden lépésben teljesül:

$$|\varepsilon_{n+1}| \le \frac{1}{2}|\varepsilon_n|,$$

ez pedig elsőrendű konvergenciát jelent. ■

Példa: $y(0)=1$, $y'=y$. Ennek a megoldása $y=e^x$ és (20.5)-ből $y_n=(1+h)^n$. Ha $n\to\infty$, akkor az analízisből tudjuk, hogy $y_n=(1+h)^n=(1+x_n/n)^n \to e^{x_n}$.

20.1.3 Definició, konvergencia

Egy numerikus módszer *lokálisan konvergens* az $x\in[0,1]$ pontban, ha $h=x/n$, $x=x_n$ mellett

$$\lim_{n\to\infty} y_n = y(x)$$

teljesül. Ha a módszer konvergens $\forall x\in[0,1]$-re, akkor azt mondjuk, a *módszer konvergens*.

20.1.4 Definició, konvergencia-sebesség

Egy konvergens numerikus módszer sebessége p-edrendű, $(1\le p)$, ha $h=x/n$, $x=x_n$ mellett $\exists M$ úgy, hogy az

$$|y(x)-y_n| \le Mh^p \tag{20.6}$$

hibabecslés teljesül, ahol M független h-tól és n-től.

20.1.5 Definició, lokális hiba v. képlethiba

Ez annak a képletnek a hibája, amellyel a következő függvényértéket közelítjük. Ilyenkor a képletbe mindenütt a pontos értéket írjuk. Például az Euler-módszernél a

$$g_{i-1} = y(x_i) - y(x_{i-1}) - hf(x_{i-1}, y(x_{i-1})), \quad i=1,2,\ldots,N \tag{20.7}$$

mennyiségek a *lokális hibák vagy képlethibák*.

20.1.6 Definició, konzisztencia

Ha $\exists p\ge 1$ és M konstans úgy, hogy

$$|g_i| \le Mh^{p+1}, \quad i=1,2,\ldots,N \tag{20.8}$$

akkor a módszer *p-edrendben konzisztens*.

A definició alapján világos, nagyobb p-re pontosabb megoldás várható.

20.1.7 Definició, globális hiba

Jelölje a pontos és numerikus megoldás eltérését

$$e_i = y(x_i) - y_i, \quad i=1,2,\ldots,N, \quad h=1/N, \quad x_i = ih. \tag{20.9}$$

Az e_i mennyiségeket *globális hibának* nevezzük.

20.1.8 Definició, stabilitás

A numerikus módszer stabil, ha van olyan K konstans, amellyel

$$|e_i| \le K\left(|e_0| + \sum_{j=1}^{i}|g_j|\right), \quad i=1,2,\ldots,N \tag{20.10}$$

teljesül, vagyis a globális hiba felülről becsülhető a lokális hibák abszolút összegével.

$$d_{n+1} \le d_n d_{n-1}, \quad n = 1,2,\ldots$$

Az indításkor $|x_0 - x^*| < 1/M$ és $|x_1 - x^*| < 1/M$, ezzel $d_0, d_1 < 1$. Igaz tehát, hogy $\exists d < 1$: $d_0, d_1 \le d$, amellyel $d_2 \le d^2$, $d_3 \le d^3$, $d_4 \le d^5$, általában

$$d_n \le d^{f_n}, \quad f_0 = f_1 = 1, \quad f_{n+1} = f_{n-1} + f_n, \quad n = 1,2,\ldots$$

Itt f_n-ek a jólismert Fibonacci-sorozat tagjai, melyeknek explicit előállítása ismert:

$$f_n = \frac{1}{\sqrt{5}}\left[b_1^{n+1} - b_2^{n+1}\right], \quad b_1 = \frac{1+\sqrt{5}}{2}, \quad b_2 = \frac{1-\sqrt{5}}{2}. \tag{16.14}$$

Mivel $|b_2| < 1$, a növekvő hatványai zérushoz fognak tartani. Emiatt létezik egy K szám, hogy minden n-re $d^{s_{n+1}} \le K$, $s_{n+1} = -b_2^{n+1}/\sqrt{5}$. Tehát írható

$$d_n \le K\left(d^{b_1/\sqrt{5}}\right)^{b_1^n} = K\left(\tilde{d}\right)^{b_1^n}, \quad \tilde{d} = d^{b_1/\sqrt{5}}.$$

Kaptuk, hogy a szelőmódszerhez tartozó hibák majorálhatók egy olyan sorozattal, amelynek konvergenciarendje $b_1 = \frac{1+\sqrt{5}}{2} \approx 1{,}62$, azaz a módszer *szuperlineáris*. ■

16.5. Példák

1. Legyen $f \in C^3[a,b]$. Parabola interpolációval készítsünk három pontra támaszkodó iterációs módszert $f(x)$ egy $[a,b]$-beli lokális minimumának meghatározására!

Megoldás. Legyen $[a,b]$-ben három pont $(x_{i-2}, f_{i-2}), (x_{i-1}, f_{i-1}), (x_i, f_i)$. Newton-interpolációval $p_2(x) = f_{i-2} + f[x_{i-2}, x_{i-1}](x - x_{i-2}) + f[x_{i-2}, x_{i-1}, x_i](x - x_{i-2})(x - x_{i-1})$. A deriváltjának zérushelye:

$$x_{i+1} = \frac{x_{i-2} + x_{i-1}}{2} - \frac{f[x_{i-2}, x_{i-1}]}{2f[x_{i-2}, x_{i-1}, x_i]}. \tag{16.15}$$

2. Egyenletes lépésközzel haladva hogyan derítenénk fel egy minimumhelyet?

Megoldás. Legyen a lépésköz h és $x_j = a + jh$, $j = 0,1,\ldots$. Az x_{j-1}, x_j, x_{j+1} alappont-hármas megfelelő, ha $f[x_{j-1}, x_j] < 0$ és $f[x_j, x_{j+1}] > 0$. Ekkor a lokális minimumot a következő egyszerűsített formulával becsülhetjük, ha (16.15)-ben x_j-t vesszük a középső pontnak:

$$x_{\min} \approx \frac{x_{j-1} + x_{j+1}}{2} - \frac{f[x_{j-1}, x_{j+1}]}{2f[x_{j-1}, x_j, x_{j+1}]} = x_j - \frac{h}{2}\frac{f_{j+1} - f_{j-1}}{f_{j+1} - 2f_j + f_{j-1}}. \tag{16.16}$$

3. A kapott iterációs módszerre fogalmazzunk meg ahhoz hasonló tételt, mint amit a Newton-módszernél láttunk a lokális konvergenciára!

Megoldás. Az egyszerűség kedvéért tekintsük (16.15)-ben azt az esetet, amikor $i = 2$. Az osztott differenciák tulajdonságait kihasználva a hibák terjedésére próbálunk egy összefüggést származtatni. Vonjuk le mindkét oldalból a minimumhelyet adó x^*-ot és legyen $\varepsilon_i = x_i - x^*$, ezzel

$$\varepsilon_3 = \frac{\varepsilon_0 + \varepsilon_1}{2} - \frac{f[x_0, x_1]}{2f[x_0, x_1, x_2]}.$$

$$|e_{i+1}| \le |e_i|(1+hL) + |g_i| \le \left(|e_{i-1}|(1+hL) + |g_{i-1}|\right)(1+hL) + |g_i| \le$$
$$\le (1+hL)^2|e_{i-1}| + (1+hL)|g_{i-1}| + |g_i| \le (1+hL)^{i+1}|e_0| + \sum_{k=0}^{i}(1+hL)^{i-k}|g_k|,$$

ebből megkapjuk a kívánt becslést, ha felhasználjuk az $(1+hL)^j \le e^{jhL} = e^{x_jL} \le e^L$ relációt:

$$|e_{i+1}| \le e^L|e_0| + \sum_{k=0}^{i} e^L|g_k| = e^L\left(|e_0| + \sum_{k=0}^{i}|g_k|\right).$$ ■

A továbbiakban rátérünk néhány fontosabb módszercsalád rövid ismertetésére.

20.3. Taylor-polinomos módszerek

Az Euler-módszer nem egyéb, minthogy vesszük a függvény elsőfokú Taylor-polinomját x_n körül és annak segítségével lépünk a következő pontba. Ebből az ötletből kiindulva készíthetünk magasabbrendű módszereket is. A következő módszert m-edrendű Taylor-polinomos módszernek nevezzük:

$$y_{n+1} = y_n + y'(x_n) + \ldots + y^{(m)}(x_n)h^m/m!, \quad n = 0,1,\ldots,N. \tag{20.11}$$

Fontos kérdés, egyáltalán kiszámíthatók-e ezek a deriváltak. Ha f elegendően sokszor differenciálható, ennek elvi akadálya nincs. Sokszor azonban súlyos gyakorlati nehézség, hogy a deriváltak nagyon hosszú, nehezen kezelhető formulákat eredményeznek. A konstrukcióból látható, m-edrendben konzisztens módszerre vezet a fenti eljárás.

20.4. Runge-Kutta módszerek

Ezek a Taylor-polinomos módszerek fenti nehézségét küszöbölik ki: nem kell magasrendű deriváltakat számolni, *a magasabb konzisztencia-rend rekurzív függvényhívásokkal is elérhető.* Az általános alak:

$$\begin{aligned} k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + ha_2, y_n + hb_{21}k_1), \\ &\vdots \\ k_j &= f(x_n + ha_j, y_n + h\sum_{l=1}^{j-1} b_{jl}k_l), \quad j = 1,2,\ldots,s, \\ y_{n+1} &= y_n + h\sum_{j=1}^{s} c_j k_j. \end{aligned} \tag{20.12}$$

Az előre kiszámolt a_i, b_{ij}, c_i paraméterek határozzák meg a konkrét módszert. Az s szám a rekurzió mélysége, más szóval: az egy lépés megtételéhez szükséges függvényhívások száma. A következő, ún. *módosított Euler-módszer* Runge-től származik:

$$\begin{aligned} k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + h/2, y_n + (h/2)k_1), \\ y_{n+1} &= y_n + hk_2. \end{aligned} \tag{20.13}$$

Deriválással megmutatható, hogy másodrendben konzisztens. Hasonlóképp másodrendű a következő Runge-Kutta módszer, amelyet az egyszerűsége miatt egy sorban írunk fel:

$$y_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_n + hf(x_n, y_n))]. \tag{20.14}$$

16.4. Newton-iteráció (Newton-Raphson módszer) és a szelőmódszer

Ha a függvény első deriváltja létezik a gyök környezetében, akkor a gyököt közelíthetjük úgy, hogy az x_n pontban a függvényhez húzott érintő metszéspontját vesszük az x tengellyel. Ez ugyanaz, mint amikor az x_n körüli elsőfokú Taylor-polinomot zérussá tesszük és x_{n+1}-re megoldjuk:

$$0 = f(x_n) + f'(x_n)(x_{n+1} - x_n),$$

innen a Newton-Raphson módszer iterációs formulája:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = F(x_n). \tag{16.7}$$

A *szelőmódszert* ebből úgy nyerjük, hogy a derivált helyére az utolsó két pontra felírt osztott differenciát tesszük:

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})} = \frac{f(x_n)x_{n-1} - f(x_{n-1})x_n}{f(x_n) - f(x_{n-1})} = F(x_{n-1}, x_n), \tag{16.8}$$

tehát ez az iterációs függvény két pontra támaszkodik. A módszer előnye a Newton-módszerrel szemben, hogy nem kell hozzá a derivált, amit néha eléggé körülményes kiszámítani. Hátránya pedig a kisebb konvergencia-sebesség.

16.4.1 Tétel, a szelőmódszer hibája

Legyen $f(x) \in C^2[x_{n-1}, x_n, x^*]$, ekkor a szelőmódszernél az $n+1$-edik iterált hibájára fennáll

$$\varepsilon_{n+1} = \varepsilon_n \varepsilon_{n-1} \frac{f''(\xi)}{2f'(\eta)}, \quad \xi, \eta \in [x^*, x_{n-1}, x_n], \tag{16.9}$$

ahol x^* a zérushely és $[x^*, x_{n-1}, x_n]$ az adott pontok által lefedett intervallum.

<u>Bizonyítás.</u> Az állítást (16.8)-ból osztott differenciák segítségével származtatjuk. Kihasználjuk, hogy $f(x^*) = 0$:

$$\varepsilon_{n+1} = x_{n+1} - x^* = x_n - x^* - (x_n - x^*)\frac{f(x_n) - f(x^*)}{x_n - x^*}\frac{1}{f[x_{n-1}, x_n]} = \varepsilon_n\left(1 - \frac{f[x^*, x_n]}{f[x_{n-1}, x_n]}\right) =$$

$$= \varepsilon_n\left(\frac{f[x_{n-1}, x_n] - f[x^*, x_n]}{f[x_{n-1}, x_n]}\frac{\varepsilon_{n-1}}{x_{n-1} - x^*}\right) = \varepsilon_n \varepsilon_{n-1}\frac{f[x^*, x_{n-1}, x_n]}{f[x_{n-1}, x_n]}$$

és innen az osztott differenciák és a deriváltak között érvényes összefüggés (14.1.1 Következmény) segítségével kapjuk az eredményt. ■

16.4.2 Következmény

A Newton-módszerre vonatkozó eredményt az $x_{n-1} \to x_n$ határátmenettel kapjuk:

$$\varepsilon_{n+1} = \varepsilon_n^2 \frac{f''(\xi)}{2f'(x_n)}, \quad \xi \in [x_n, x^*], \tag{16.10}$$

Látjuk, ha van konvergencia, akkor az másodrendű, feltéve, hogy $f'(x^*) \neq 0$.

fixpont egyenletet. $F(y)$ az általános feltételünk alapján eleget tesz a Lipschitz-feltételnek $hL/2$ állandóval, így $F(y)$ kontrakció, ha $h < 2/L$. Célszerű formája az iterációnak:

$$\begin{aligned} y_{n+1}^0 &= y_n + hf_n, && \text{kezdőérték az Euler-módszerből} \\ y_{n+1}^{k+1} &= y_n + \frac{h}{2}\left(f_n + f(x_{n+1}, y_{n+1}^k)\right), && k \geq 0. \end{aligned} \tag{20.20}$$

Leállás: ha $\left|y_{n+1}^{k+1} - y_{n+1}^k\right| < \varepsilon(1-q)$, ahol q a konvergencia-tényező és ε a kívánt hibakorlát.

A trapézmódszer másodrendben konzisztens és stabil, ha $f \in C^2([0,1] \times \mathbb{R})$ és $h < 1/L$, tehát ekkor másodrendű konvergenciára számíthatunk.

További többlépéses módszereket a zárt vagy a jobbról nyilt kvadratúra-formulák segítségével lehet származtatni.

20.6. Aszimptotikus stabilitás

A gyakorlati számítások szempontjából nem elegendő, ha egy módszer konzisztens és stabil. A kezdeti hiba (- ha y_0 is hibával terhelt, -) továbbterjedése szempontjából fontos egy további stabilitási tulajdonság. Egy differenciálegyenlet-megoldó numerikus módszer *aszimptotikusan stabil*, ha az

$$y(0) = 1, \quad y'(x) = q\,y(x), \quad q < 0 \tag{20.21}$$

tesztfeladatra alkalmazva a kapott numerikus közelítések sorozatára minden $h > 0$ lépésköz esetén fennáll $y_n \to 0$, ha $n \to \infty$. Ezt a stabilitási fogalmat a szakirodalom gyakran A_0-stabilitásnak nevezi.

A tesztfeladat megoldása x növekedésével zérushoz tart: $y(x) = e^{qx} \to 0$, $x \to \infty$, mivel q negatív. Így a definíció azt követeli a módszertől, hogy a lépésköztől függetlenül a numerikus megoldás is tartsa meg ezt a lecsengő jelleget. Kimutatható, az Euler-módszer nem aszimptotikusan stabil, de a trapézmódszer az.

$$|F(x)-F(y)| \le \max_{x\in S}|F'(x)||x-y|, \quad x,y\in S$$

tehát F kontrakció $q=\max_{x\in S}|F'(x)|<1$ kontrakciós állandóval. ■

16.2.3 Következmény

Ha $F(x)$ kontrakció, akkor a Banach fixponttétel szerint csak egy gyök van és a kontrakciós állandó ismeretében a közelítés pontosságát is becsülni tudjuk.

Visszatérve a fenti példához: a kapott iteráció biztosan konvergens abban a tartományban, ahol $\left(\sqrt{\sin(x)}\right)' = \dfrac{\cos(x)}{2\sqrt{\sin(x)}} < 1$. Látjuk, ha $x=0$ vagy $x=\pi$, ezzel a kifejezéssel baj van, mert a formula kiértékelésekor 0-val kéne osztani. De például $x=\pi/4$ esetén a kifejezés értéke $1/\sqrt{8}$, ami már jobbnak tűnik. Ha megrajzoljuk az x^2 parabolát és a $\sin(x)$ függvény képét, látjuk, hogy két nemnegatív gyök van, az egyik a zérus, a másik pedig közel $x=\pi/4$-hez, tehát remélhető, hogy az iteráció $\pi/4$-gyel indítva konvergens. De az is látszik, hogyha nagyon kicsi pozitív értékkel indítjuk az iterációt, akkor sem kapjuk meg a zérus gyököt, mert az iteráció mindig elvisz a nagyobbik gyök irányába.

Ha azonban az $x=\arcsin(x^2)$ iterációt készítjük, könnyen meggyőződhetünk arról, hogy kis pozitív x-re zérushoz tart. Ha azonban $x=1$-gyel indítunk, akkor először a $\pi/2$ értéket kapjuk, majd komplex számokat, mivel az argumentum nagyobb 1-nél.

A tanulság: ügyelnünk kell, a kapott függvény hova képez le, és a leképezés tartományában megmaradnak-e a konvergencia tulajdonságok, illetve azt a gyököt kapjuk-e, amit szeretnénk meghatározni.

Ha a megoldandó egyenletben több helyen is szerepel x, akkor több $x=F(x)$ kifejezés is készíthető. Például szerepeljen két helyen, ekkor meg lehet mutatni: a kapott két iteráció egy adott helyen egyszerre nem lehet konvergens. Legyen ugyanis $f(x_1,x_2)=0$ a megoldandó egyenlet, ahol a két előfordulást x_1 és x_2-vel azonosítjuk. Legyen $F_i(x)$ az az iterációs függvény, amelyet x_i kifejezésével kaptunk. Ez azt jelenti, hogy

$$f(F_1(x),x)=0 \quad \text{és} \quad f(x,F_2(x))=0. \tag{16.3}$$

Legyen α egyszeres gyök: $f(\alpha,\alpha)=0$. Ha a (16.3)-ben szereplő kifejezéseket deriváljuk x szerint és helyettesítjük $x=\alpha$-t, az eredmény:

$$f_1'(\alpha,\alpha)F_1'(\alpha)+f_2'(\alpha,a)=0,$$
$$f_1'(\alpha,\alpha)+f_2'(\alpha,\alpha)F_2'(\alpha)=0,$$

ahol f alsó indexe azt jelöli, melyik hely szerint deriváltunk. Ahhoz, hogy egyszeres gyök mellett nemzérus megoldást kapjunk $f_i'(\alpha,\alpha)$-ra kell, hogy a kapott rendszer determinánsa 0 legyen:

$$\begin{vmatrix} F_1'(\alpha) & 1 \\ 1 & F_2'(\alpha) \end{vmatrix} = F_1'(\alpha)F_2'(\alpha)-1=0,$$

ahonnan

$$\left|F_1'(\alpha)\right| = 1/\left|F_2'(\alpha)\right|. \tag{16.4}$$

Hacsak nem 1 abszolút értékűek a deriváltak, a gyök közelében az egyik iterációs függvény konvergens, a másik meg divergens lesz. Hasonlóan lehet vizsgálni azt az esetet, amikor x 2-nél