3316321 Numeerinen analyysi

Timo Erkama
LaTeX-käännös
Marko Lamminsalo
Itä-Suomen yliopisto
20. joulukuuta 2010

Sisältö

1 Virheanalyysi

1.1 Virheet numeerisissa laskuissa

Käytännön laskuissa esiintyy vain äärellisen monta desimaalia ja laskutoimitusta; tulokset ovat siten yleensä *likiarvoja*.

Jos suureen tarkka arvo on a ja likiarvo $\tilde{a}$, niin erotus

$$\varepsilon = \tilde{a} - a$$

on (absoluuttinen) *virhe* likiarvossa $\tilde{a}$. Toisin sanoen

$$\tilde{a} = a + \varepsilon,$$

eli likiarvo on tarkan arvon ja virheen summa.

Esimerkki.

$$\begin{array}{ll} \tilde{a} = 10.5, \quad a = 10.2 & \Rightarrow \varepsilon = 0.3 \\ \tilde{a} = 1.60, \quad a = 1.82 & \Rightarrow \varepsilon = -0.22 \end{array}$$

Likiarvon $\tilde{a}$ *suhteellinen virhe* ε_r on

$$\varepsilon_r = \frac{\varepsilon}{a} = \frac{\tilde{a} - a}{a} = \frac{\text{virhe}}{\text{tarkka arvo}} \quad (a \neq 0)$$

Selvästi

$$\varepsilon_r \approx \frac{\varepsilon}{\tilde{a}}$$

jos $|\varepsilon|$ on paljon pienempi kuin $|\tilde{a}|$. Virheen vastaluku $\gamma = a - \tilde{a} = -\varepsilon$ on *korjaus*, jolloin tarkka arvo saadaan likiarvon ja korjauksen summana

$$a = \tilde{a} + \gamma.$$

Virheen yläraja on luku β siten, että

$$|\tilde{a} - a| \leq \beta, \quad \text{siis } |\varepsilon| \leq \beta.$$

Virhelähteet

1. Idealisoinnit matemaattisessa mallissa

2. Lähtöarvovirheet

3. Katkaisuvirheet (äärellisten tai äärettömien prosessien ennenaikaisesta keskeyttämisestä aiheutuvat virheet tai muut yksinkertaistukset numeerisessa mallissa)

4. Pyöristysvirheet

Lisäksi ajatus-, kirjoitus-, lasku- ym. huolimattomuusvirheet.

1.2 Liukuluvut

Reaaliluvun desimaaliesityksessä tarvitaan yleensä äärettömän monta numeroa. Laskimet ja tietokoneet kykenevät kuitenkin käsittelemään vain äärellisiä merkkijonoja. Sen vuoksi reaaliluvut on laskutoimituksia varten korvattava ns. *liukuluvuilla*, jolloin merkitsevien numeroiden lukumäärä kussakin luvussa on jokin koneesta riippuva vakio. Tätä varten on kukin reaaliluku ensin *pyöristettävä* käyttämällä seuraavia pyöristyssääntöjä:

Luku pyöristetään n:ään desimaaliin siten, että

- poistetaan kaikki n:nnen oikealla puolella olevat desimaalit
- jos poistettu luku on $> \frac{1}{2} \cdot 10^{-n}$, korotetaan n:s desimaali yhdellä yksiköllä
- jos poistettu luku on $= \frac{1}{2} \cdot 10^{-n}$, korotetaan n:s desimaali vain mikäli se on *pariton*

Esimerkki. Pyöristetään muutamia lukuja kolmeen desimaaliin.

$$
\begin{aligned}
0.4711 &\approx 0.471 \\
0.4716 &\approx 0.472 \\
0.4715 &\approx 0.472 \\
0.4705 &\approx 0.470
\end{aligned}
$$

Viimeinen sääntö johtuu siitä, että pyritään eliminoimaan systemaattinen pyöristysvirhe. Desimaalin korotus tapahtuu "yhtä usein" kuin korottamatta jättäminen.

Pyöristettäessä n:ään desimaaliin syntyvä virhe on ns. *pyöristysvirhe* on itseisarvoltaan enintään $\frac{1}{2} \cdot 10^{-n}$.

Äskeisen esimerkin mukaan pyöristetyissä luvuissa on kolme merkitsevää numeroa. Sanomme, että jonkin reaaliluvun likiarvossa on n *merkitsevää numeroa*, jos n on suurin luonnollinen luku siten, että absoluuttinen virhe on itseisarvoltaan korkeintaan ensimmäisen nollasta eroavan numeron yksikkö kerrottuna luvulla $5 \cdot 10^{-n}$.

Jos esimerkiksi tarkka arvo $a = 1.1996$ ja likiarvo $\tilde{a} = 1.200$, niin likiarvossa $\tilde{a}$ on neljä merkitsevää numeroa.

Vastaavasti sanotaan,ettö likiarvossa on n *oikeaa desimaalia*, jos absoluuttinen virhe on enintään $\frac{1}{2} \cdot 10^{-n}$.

Merkitsevien numeroiden lukumäärä kuvaa suhteellista tarkkuutta ja oikeiden desimaalien lukumäärän absoluuttista tarkkuutta.

Esimerkki.

(a)	$\tilde{a} = 47.11$,	$\lvert\varepsilon\rvert \leq 0.5 \cdot 10^{-2}$	2 oik. des., 4 merk. nroa
(b)	$\tilde{a} = 0.0047110$,	$\lvert\varepsilon\rvert \leq 0.5 \cdot 10^{-7}$	7 oik. des., 5 merk. nroa
(c)	$\tilde{a} = 4710 \cdot 10^{2}$,	$\lvert\varepsilon\rvert \leq 0.5 \cdot 10^{2}$	0 oik. des., 4 merk. nroa
(d)	$\tilde{a} = 47100$,	$\lvert\varepsilon\rvert \leq 0.5 \cdot 10^{2}$	0 oik. des., 3 merk. nroa

Merkitsevien numeroiden lukumäärä saadaan seuraavista yhtälöistä:

$$\begin{array}{lll}
\text{(a)} & 0.5 \cdot 10^{-2} = 10 \cdot 5 \cdot 10^{-n} & \Leftrightarrow \quad n = 4 \\
\text{(b)} & 0.5 \cdot 10^{-7} = 10^{-3} \cdot 5 \cdot 10^{-n} & \Leftrightarrow \quad n = 5 \\
\text{(c)} & 0.5 \cdot 10^{2} = 10^{5} \cdot 5 \cdot 10^{-n} & \Leftrightarrow \quad n = 4
\end{array}$$

Reaalilukujen desimaaliesityksessä kantalukuna on 10. Tietokoneiden lukujärjestelmän kantalukuna on usein jokin muu luku $\beta \geq 2$, esimerkiksi $\beta = 2$ tai $\beta = 16$. Tällaisessa lukujärjestelmässä jokaista reaalilukua vastaa päättymätön merkkijono

$$(\pm d_n d_{n-1} \ldots d_2 d_1 d_0 . d_{-1} d_{-2} \ldots)_\beta$$

missä $d_n, d_{n-1}, \ldots$ ovat lukujen 0 ja $\beta - 1$ välillä sijaitsevia kokonaislukuja. Merkkijonoa vastaava reaaliluku on

$$d_n \beta^n + d_{n-1} \beta^{n-1} + \cdots + d_2 \beta^2 + d_1 \beta^1 + d_0 \beta^0 + d_{-1} \beta^{-1} + d_{-2} \beta^{-2} + \cdots .$$

Esimerkiksi $\pi = 3.1415 \ldots = 3 \cdot 10^0 + 1 \cdot 10^{-1} + 4 \cdot 10^{-2} + 1 \cdot 10^{-3} + \cdots$.

Esimerkki.

$$
\begin{aligned}
(760)_8 &= 7 \cdot 8^2 + 6 \cdot 8^1 + 0 \cdot 8^0 = (496)_{10} \\
(101.101)_2 &= 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3} \\
&= (5.625)_{10} \\
(0.333)_{10} &= 3 \cdot 10^{-1} + 3 \cdot 10^{-2} + 3 \cdot 10^{-3} + \cdots = \frac{1}{3} \\
\frac{1}{5} &= (0.2)_{10} = (0.00110011\ldots)_2 \\
\frac{1}{5} &= 2^q + r
\end{aligned}
$$

Tietokoneissa reaaliluvut korvataan ns. liukuluvuilla, jolloin kullekin luvulle varataan kiinteä määrä muistitilaa. Jos lukujärjestelmän kantaluku on β, jokainen nollasta eroava reaaliluku voidaan esittää muodossa

$$X = M \cdot \beta^e,$$

missä e on kokonaisluku ja

$$M = \pm D_0.D_1D_2D_3\ldots$$

$$0 \leq D_i \leq \beta - 1$$

$$D_0 \neq 0.$$

Liukulukuesityksessä M pyöristetään äärelliseksi merkkijonoksi

$$m = \pm d_0.d_1d_2\ldots d_t,$$

jossa on kiinteä määrä $(t+1)$ numeroita. Tietokoneen muistiin varastoidaan tällöin luku

$$x = m \cdot \beta^e$$

(tai $x = \beta^{e+1}$, mikäli esimerkiksi $D_0 = D_1 = D_2 = \ldots = \beta - 1$). Tässä m on ns. *jaososa* eli mantissa ja e *eksponenttiosa*. Koska $0 \leq d_i \leq \beta - 1$ ja $d_0 \neq 0$, jokainen nollasta eroava liukuluku on normalisoitu siten, että $1 \leq |m| < \beta$.

Eksponentille e varatusta muistitilasta riippuu kuinka suuria tai pieniä lukuja kone kykenee käsittelemään. Eksponentille e asetettava rajoitus on muotoa

$$L \leq e \leq U$$

missä L on negatiivinen ja U positiivinen kokonaisluku. Jos jonkin laskutoimituksen tuloksena saatavan liukuluvun eksponentille pätee $e > U$, syntyy ns. *ylivuoto* ja kone antaa virheilmoituksen. *Alivuoto* $e < L$ ei yleensä aiheuta prosessin keskeytymistä.

1.3 Numeerinen stabiilisuus

Numeeriseen probleemaan liittyvä *algoritmi* on täydellinen kuvaus niistä äärellisen monesta laskutoimituksesta, joiden välityksellä probleeman lähtöarvoista saadaan selville tulokset eli probleeman ratkaisu.

Algoritmi on *stabiili*, jos siihen liittyvillä katkaisu- ja pyöristysvirheillä on vain pieni vaikutus lopputuloksiin. Päinvastaisessa tapauksessa algoritmi on *epästabiili*. Tällainen *numeerinen epästabiilisuus* voidaan usein välttää valitsemalla parempi algoritmi.

Niin sanotun *matemaattisen epästabiilisuuden* tapauksessa tilanne on toinen; probleemaa kutsutaan tällöin *pahanlaatuiseksi* tai *häiriöalttiiksi* ja lopputulos voi olla epätarkka riippumatta käytetystä algoritmista.

Esimerkki. Etsi yhtälöiden

$$\text{(a) } x^2 - 4x + 2 = 0 \quad \text{ja} \quad \text{(b) } x^2 - 40x + 2 = 0 \tag{1}$$

juuret käyttämällä laskutoimituksissa neljää merkitsevää numeroa.

Toisen asteen yhtälön $ax^2 + bx + c = 0$ juuret x_1 ja x_2 saadaan kaavoista

$$x_1 = \frac{1}{2a}\left(-b + \sqrt{b^2 - 4ac}\right), \quad x_2 = \frac{1}{2a}\left(-b - \sqrt{b^2 - 4ac}\right). \tag{2}$$

Koska $x_1 x_2 = \frac{c}{a}$, vaihtoehtoiset kaavat ovat

$$x_1 \text{ kuten edellä}, \quad x_2 = \frac{c}{a x_1}. \tag{3}$$

Yhtälölle (1a) saadaan kaavoista (2)

$$x = 2 \pm \sqrt{2} = 2.000 \pm 1.414 \;\Rightarrow\; \begin{cases} x_1 = 3.414 \\ x_2 = 0.586 \end{cases}$$

ja kaavoista (3)

$$x_1 = 3.414, \quad x_2 = \frac{2.000}{3.414} = 0.5858.$$

Virhe jälkimmäisessä x_2:n arvossa on $\leq 10^{-4}$.

Yhtälölle (1b) saadaan kaavoista (2)

$$x = 20 \pm \sqrt{398} = 20.00 \pm 19.95 \ \Rightarrow \ \begin{cases} x_1 = 39.95 \\ x_2 = 0.05 \end{cases}$$

ja kaavoista (3):

$$x_1 = 39.95, \quad x_2 = \frac{2.000}{39.95} = 0.05006.$$

Ensimmäisessä likiarvossa $x_2 = 0.05$ on vain yksi merkitsevä numero, mutta jälkimmäisessä niitä on 4. Virhe jälkimmäisessä x_2:n arvossa on $\leq 10^{-5}$.

1.4 Differentiaalilaskennan käyttö virheanalyysissä

Lasketaan funktion $f(x) = \frac{1}{x^2}$ arvo pisteessä $x = 0.015$, mutta pyöristetään ensin x kahteen desimaaliin. Tällöin syntyy virhe

$$f(0.02) - f(0.015) = \frac{1}{(0.02)^2} - \frac{1}{(0.015)^2} = -1944.$$

Pieni häiriö (0.005) x:n arvossa aiheuttaa suuren virheen. Tehtävä on *häiriöaltis.*

Lopputuloksen virheen suhde lähtöarvon virheeseen (virheen suurennussuhde) on tässä tapauksessa

$$-\frac{1944}{0.005} \approx -388000.$$

Tutkimalla vastaavaa osamäärää eri lähtöarvoilla saadaan käsitys tehtävän *häiriöalttiudesta.*

Jos f on jatkuvasti derivoituva funktio, niin väliarvolauseen nojalla

$$f(x + \varepsilon) - f(x) = f'(\xi)\varepsilon,$$

missä ξ on lukujen x ja $x+\varepsilon$ välissä.

$$\frac{|f(x+\varepsilon)-f(x)|}{|\varepsilon|} \leq \max|f'(\xi)|$$

$$|f(x+\varepsilon)-f(x)| \leq |\varepsilon| \max|f'(\xi)| \tag{1}$$

Virheen suurennussuhde laskettaessa arvoa $f(x)$ on siis korkeintaan f:n derivaatan itseisarvon maksimi pisteiden x ja $x+\varepsilon$ välissä. Samoin nähdään, että suurennussuhde on itseisarvoltaan vähintään $\min|f'(\xi)|$.

Useamman muuttujan funktioille voidaan johtaa vastaava tulos. Merkitään $f = f(x_1, x_2, \ldots, x_n)$ funktion tarkkaa arvoa pisteessä $(x_1, x_2, \ldots, x_n)$ ja olkoon $\tilde{f} = f(x_1+\varepsilon_1, x_2+\varepsilon_2, \ldots, x_n+\varepsilon_n)$ vastaava likiarvo. Silloin

$$\tilde{f} - f = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x+\theta\varepsilon)\varepsilon_k,$$

missä $0 < \theta < 1$, $x = (x_1, \ldots, x_n)$ ja $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$.

$$\Rightarrow \quad |\tilde{f} - f| \leq \sum_{k=1}^{n} |\varepsilon_k| \max\left|\frac{\partial f}{\partial x_k}\right|, \tag{2}$$

missä maksimit lasketaan pisteitä $x = (x_1, \ldots, x_n)$ ja $x+\varepsilon = (x_1+\varepsilon_1, \ldots, x_n+\varepsilon_n)$ yhdistävällä janalla.

Esimerkki 1. Arvioi ylöspäin virhettä, joka syntyy laskettaessa funktion

$$f(x_1, x_2, x_3) = \frac{x_1}{x_2^2 + x_3^2}$$

arvoa pisteessä (1.0,1.0,1.0), missä koordinaatit on annettu kahden merkitsevän numeron tarkkuudella.

Virhe ε_k koordinaatin x_k arvossa on itseisarvoltaan $\leq 0.05 \quad (1 \leq k \leq 3)$.

$$\frac{\partial f}{\partial x_1} = \frac{1}{x_2^2+x_3^2}, \quad \frac{\partial f}{\partial x_2} = \frac{-2x_2x_1}{(x_2^2+x_3^2)^2}, \quad \frac{\partial f}{\partial x_3} = \frac{-2x_3x_1}{(x_2^2+x_3^2)^2}$$

$$\max \left|\frac{\partial f}{\partial x_1}\right| \leq \frac{1}{0.95^2 + 0.95^2} = 0.56$$
$$\max \left|\frac{\partial f}{\partial x_2}\right| \leq \frac{2 \cdot 1.05 \cdot 1.05}{(0.95^2 + 0.95^2)^2} = 0.68$$
$$\max \left|\frac{\partial f}{\partial x_3}\right| \leq \frac{2 \cdot 1.05 \cdot 1.05}{(0.95^2 + 0.95^2)^2} = 0.68$$

Epäyhtälön (2) nojalla

$$\begin{aligned}|\tilde{f} - f| &= |f(1+\varepsilon_1, 1+\varepsilon_2, 1+\varepsilon_3) - f(1,1,1)| \\ &\leq 0.05 \cdot 0.56 + 0.05 \cdot 0.68 + 0.05 \cdot 0.68 = 0.096.\end{aligned}$$

Esimerkki 2. $f = x_1 + x_2 + \ldots + x_n$

$$\frac{\partial f}{\partial x_k} = 1$$

$$(2) \Rightarrow |\tilde{f} - f| \leq \sum_{k=1}^{n} |\varepsilon_k|$$

Yhteen- ja vähennyslaskussa absoluuttisten virheiden ylärajat lasketaan yhteen.

Oletetaan, että $n = 1000$ ja $|\varepsilon_k| \leq 0.5 \cdot 10^{-5}$. Kaavan (2) nojalla

$$|\tilde{f} - f| \leq 1000 \cdot 0.5 \cdot 10^{-5} = 0.5 \cdot 10^{-2}.$$

Todellinen virhe on todennäköisesti paljon pienempi. Yläraja $0.5 \cdot 10^{-2}$ saavutetaan vain jos kaikki ε_k:t ovat samanmerkkisiä ja itseisarvoltaan $= 0.5 \cdot 10^{-5}$.

Esimerkki 3. $f = x_1 - x_2$

$$\left|\frac{\partial f}{\partial x_1}\right| = \left|\frac{\partial f}{\partial x_1}\right| = 1$$

Absoluuttinen virhe:

$$|\tilde{f} - f| \leq |\varepsilon_1| + |\varepsilon_2|$$

Suhteellinen virhe:

$$\frac{|\tilde{f} - f|}{|f|} \leq \frac{|\varepsilon_1| + |\varepsilon_2|}{|x_1 - x_2|}$$

Jos $x_1 = 0.5763 \pm 0.5 \cdot 10^{-4}$ ja $x_2 = 0.5765 \pm 0.5 \cdot 10^{-4}$, saadaan

$$\frac{|\tilde{f} - f|}{|f|} \leq \frac{10^{-4}}{10^{-4}} = 1 = 100\%.$$

Kun vähennettävä ja vähentäjä ovat likimain samat, voi syntyä suuri suhteellinen virhe. Tämä voidaan usein välttää käyttämällä sopivaa kiertotietä.

Esimerkki 4.

$$f = x_1^{m_1} \cdot x_2^{m_2} \cdots x_n^{m_n} = \left(\frac{f}{x_k^{m_k}}\right) x_k^{m_k}$$

$$\left|\frac{\partial f}{\partial x_k}\right| = \left(\frac{f}{x_k^{m_k}}\right) m_k x_k^{m_k - 1} = \frac{m_k}{x_k} f$$

Absoluuttisen virheen yläraja epäyhtälön (2) nojalla:

$$|\tilde{f} - f| \leq \sum_{k=1}^{n} |\varepsilon_k| \max \left|\frac{m_k}{x_k} f\right|$$

Suhteellisen virheen yläraja:

$$\frac{|\tilde{f} - f|}{|f|} \leq \sum_{k=1}^{n} |\varepsilon_k||m_k| \frac{\max |f/x_k|}{|f|} \approx \sum_{k=1}^{n} |m_k| |\frac{\varepsilon_k}{x_k}|$$

Kerto- ja jakolaskussa suhteellisten virheiden ylärajat lasketaan yhteen.

Esimerkki 5. Olkoot x ja y lukuja siten, että $5 \leq x \leq 10$ ja $1 \leq y \leq 2$. Oletetaan, että x on annettu kolmen ja y neljän oikean desimaalin tarkkuudella. Arvioi absoluuttista virhettä luvussa $e^{\sin(xy)}$.

$$f(x, y) = e^{\sin(xy)}$$

$$(2) \Rightarrow |\tilde{f} - f| \leq |\varepsilon_x| \max \left|\frac{\partial f}{\partial x}\right| + |\varepsilon_y| \max \left|\frac{\partial f}{\partial y}\right|$$

Tässä $|\varepsilon_x| \leq \frac{1}{2} \cdot 10^{-3}$ ja $|\varepsilon_y| \leq \frac{1}{2} \cdot 10^{-4}$.

$$\frac{\partial f}{\partial x} = y \cos(xy) e^{\sin(xy)}; \qquad \frac{\partial f}{\partial y} = x \cos(xy) e^{\sin(xy)}$$

Kun $5 \leq x \leq 10$ ja $1 \leq y \leq 2$, niin

$$\left|\frac{\partial f}{\partial x}\right| \leq 2 \cdot 1 \cdot e^1 = 2e \quad \text{ja} \quad \left|\frac{\partial f}{\partial y}\right| \leq 10 \cdot 1 \cdot e^1 = 10e,$$

joten

$$|\tilde{f} - f| \leq \frac{1}{2} \cdot 10^{-3} \cdot 2e + \frac{1}{2} \cdot 10^{-4} \cdot 10e = 1.5e \cdot 10^{-3} < 0.005.$$

Esimerkki 6. Kuinka monta termiä saa esiintyä summassa

$$\sum_{k=1}^{N} \frac{1}{\sqrt{1 + cosx_k}}$$

kun vaaditaan, että absoluuttisen virheen itseisarvo on $\leq 10^{-2}$ ja oletetaan, että luvut x_k on annettu kolmen desimaalin tarkkuudella ja $0 \leq x_k \leq 1$ kaikille k. Muita virhelähteitä ei oteta huomioon.

Yksittäisen termin aiheuttama virhe on enintään

$$|\varepsilon| \max |f'(\xi)|,$$

missä

$$f(x) = \frac{1}{\sqrt{1 + cosx_k}}, \quad 0 \leq \xi \leq 1, \quad |\varepsilon| \leq \frac{1}{2} \cdot 10^{-3}.$$

Koska

$$|f'(x)| = \left| \frac{\sin x}{2(1 + \cos x)^{\frac{3}{2}}} \right|$$

on kasvava välillä $0 \leq x \leq 1$,

$$\max_{0 \leq \xi \leq 1} |f'(\xi)| \leq \frac{\sin 1}{2(1 + \cos 1)^{\frac{3}{2}}} \approx 0.2201.$$

Yksittäisen termin virhe on siis $\leq \frac{1}{2} \cdot 10^{-3} \cdot 0.2201$. Summan virheen yläraja saadaan kertomalla tässä luvulla N. Näin ollen N voi olla korkeintaan niin suuri, että epäyhtälö

$$N \cdot \frac{1}{2} \cdot 10^{-3} \cdot 0.2201 \leq 10^{-2}$$

pätee. Ratkaisemalla tästä N saadaan

$$N \leq \frac{2 \cdot 10^{-2}}{10^{-3} \cdot 0.2201} = 90.9.$$

Suurin lukua 90.9 pienempi kokonaisluku on 90, joten vastaus on $N = 90$.

1.5 Tilastollinen virheanalyysi

Jos laskutoimituksia on paljon kuten esimerkeissä 2 ja 6, ovat yo. arviot virheen ylärajalle yleensä liian pessimistisiä. Realistisempi arvio saadaan toisinaan käyttämällä tilastollista analyysiä.

Esimerkiksi jos kerrotaan keskenään 1000 lukua, joissa on kolme merkitsevää numeroa, niin kussakin luvussa suhteellinen virhe on enintään 0.5 ‰ mutta tulon suhteellinen virhe on ≤ 50 %. Tilastollisella analyysillä nähdään kuitenkin, että 68 prosentin todennäköisyydellä virhe on $< 3,2$ %.

2 Numeerinen integrointi

2.1 Puolisuunnikassääntö ja Simpsonin sääntö

Haluamme laskea numeerisesti integraalin

$$\int_a^b f(x)dx.$$

Jos integroitava funktio f on jonkin tunnetun funktion F derivaatta, ts. $f(x) = F'(x)$, saadaan *analyyttinen ratkaisu*

$$F(b) - F(a) = \int_a^b F'(x)dx = \int_a^b f(x)dx. \tag{1}$$

Vaikka F tunnettaisiinkin, niin oikea puoli voi olla niin hankala laskea, että on parempi turvautua johonkin *numeeriseen ratkaisumenetelmään.*

On laskettava likimääräisesti viivoitetun alueen pinta-ala. Jaetaan alue vyöhykkeisiin kuten kuvassa. Approksimoidaan kunkin vyöhykkeen pinta-alaa ja lasketaan likiarvot yhteen, jolloin integraalin $\int_a^b f(x)dx$ arvolle saadaan arvio

$$\int_a^b f(x)dx = \int_a^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \cdots + \int_{x_{n-1}}^{b} f(x)dx. \tag{2}$$

Jos merkitään $x_0 = a$ ja $x_n = b$, niin (2) voidaan kirjoittaa

$$\int_a^b f(x)dx = \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} f(x)dx. \tag{3}$$

Oletetaan, että jakopisteet x_k on valittu siten, että kaikilla osaväleillä $[x_k, x_{k+1}]$ on sama pituus h. Approksimoidaan osaväliä $[x_k, x_{k+1}]$ vastaavan vyöhykkeen pinta-alaa kuvaan piirretyn puolisuunnikkaan pinta-alalla. Tällöin saadaan likiarvo

$$\int_{x_k}^{x_{k+1}} f(x)dx \approx h\frac{f(x_k) + f(x_{k+1})}{2}.$$

Sijoittamalla tämä kaavaan (3) saadaan

$$\int_a^b f(x)dx \approx \sum_{k=0}^{n-1} h \frac{f(x_k) + f(x_{k+1})}{2}. \tag{4}$$

Funktion f kuvaajaa on approksimoitu *murtoviivalla.*

Kaavassa (4) jokainen arvo $f(x_k)$ esiintyy kahdesti (paitsi $f(x_0)$ ja $f(x_n)$). Näin ollen (4) voidaan kirjoittaa

$$\int_a^b f(x)dx \approx h \left[\frac{f(x_0)}{2} + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{f(x_n)}{2} \right] \tag{5}$$

Tämä on ns. *puolisuunnikassääntö.*

Jos f on kaksi kertaa jatkuvasti derivoituva välillä $[a, b]$, niin voidaan näyttää, että virhe approksimoitaessa on muotoa

$$-\frac{(b-a)}{12} h^2 f''(\xi),$$

missä ξ on jokin piste välillä $[a, b]$ ($a = x_0$, $b = x_n$).

Puolisuunnikassääntö

$$\int_a^b f(x)dx = h \left[\frac{f(x_0)}{2} + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{f(x_n)}{2} \right] + R, \tag{6}$$

missä

$$R = -\frac{(b-a)}{12} h^2 f''(\xi), \quad a = x_0 \le \xi \le x_n = b.$$

Jäännöstermille R saadaan arvio

$$|R| \le \frac{(b-a)}{12} h^2 \max_{a \le \xi \le b} |f''(\xi)| \tag{7}$$

Virheen yläraja saadaan mielivaltaisen pieneksi lisäämällä osavälien lukumäärää, sillä $R \to 0$ kun $h \to 0$.

Esimerkki 1. Sovelletaan puolisuunnikassääntöä (6) funktioon $f(x) = x^2$ välillä $[0, 1]$. Valitaan

$$h = \frac{1}{n}, \quad x_0 = 0, \quad x_n = 1, \quad x_k = kh = \frac{k}{n}.$$

Silloin $f'(x) = 2x$ ja $f''(x) = 2$, joten

$$R = -\frac{1-0}{12}\left(\frac{1}{n}\right)^2 \cdot 2 = -\frac{1}{6n^2}$$

$$\begin{aligned}
\frac{1}{3} = \int_0^1 x^2 dx &= h\left[\frac{x_0^2}{2} + x_1^2 + x_2^2 + \cdots + x_{n-1}^2 + \frac{x_n^2}{2}\right] + R \\
&= h\left[\frac{0}{2} + \left(\frac{1}{n}\right)^2 + \left(\frac{2}{n}\right)^2 + \cdots + \left(\frac{n-1}{n}\right)^2 + \frac{1}{2}\right] - \frac{1}{6n^2} \\
&= \frac{h}{n^2}\left[1^2 + 2^2 + \cdots + (n-1)^2\right] + \frac{h}{2} - \frac{1}{6n^2}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow 1^2 + 2^2 + \cdots + (n-1)^2 &= \frac{n^2}{h}\left[\frac{1}{3} - \frac{h}{2} + \frac{1}{6n^2}\right] \\
&= \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6} \\
&= \frac{2n^3 - 3n^2 + n}{6}
\end{aligned}$$

$$\begin{aligned}
1^2 + 2^2 + \cdots + (n-1)^2 + n^2 &= \frac{2n^3 - 3n^2 + n}{6} + n^2 \\
&= \frac{2n^3 + 3n^2 + n}{6}
\end{aligned}$$

Jos $\min_{x\in[a,b]} f''(x) << \max_{x\in[a,b]} f''(x)$, virhearvio (7) on yleensä karkea. Jos taas $\min_{x\in[a,b]} f''(x) \approx \max_{x\in[a,b]} f''(x)$, voidaan virhettä arvioida tarkemmin käyttämällä ns. *Richardsonin ekstrapolointia* seuraavasti:

Käytetään puolisuunnikassääntöä ensin siten, että osavälien lukumäärä on n ($h = \frac{b-a}{n}$); saadaan

$$\int_a^b f(x)dx = T_1 + R_1, \quad \text{missä } R_1 = -\frac{(b-a)}{12}\left(\frac{b-a}{n}\right)^2 f''(\xi_1). \tag{8}$$

Sitten puolitetaan jokainen osaväli ja sovelletaan puolisuunnikassääntöä siten, että osavälien lukumäärä on $2n$. Tällöin jokaisen osavälin pituus on $\frac{h}{2}$, ja saadaan

$$\int_a^b f(x)dx = T_2 + R_2, \quad \text{missä } R_2 = -\frac{(b-a)}{12}\left(\frac{b-a}{2n}\right)^2 f''(\xi_2). \tag{9}$$

Jos $\min_{x\in[a,b]} f''(x) \approx \max_{x\in[a,b]} f''(x)$, niin $f(\xi_1) \approx f(\xi_2)$. Tällöin $R_1 \approx 4R_2$, joten yhtälöiden (8) ja (9) nojalla saadaan

$$\begin{cases} \int_a^b f(x)dx \approx T_1 + 4R_2 \\ \int_a^b f(x)dx = T_2 + R_2 \end{cases}$$

Vähentämällä saadaan *kolmasosasääntö*

$$R_2 \approx \frac{T_2 - T_1}{3}. \tag{10}$$

Tämä on arvio puolisuunnikassäännön jäännöstermille kun osavälejä on $2n$ kappaletta.

Koska $\int_a^b f(x)dx = T_2 + R_2$, niin kaavan (10) avulla voidaan *parantaa* puolisuunnikassäännön antamaa likiarvoa T_2 integraalille $\int_a^b f(x)dx$. Saadaan

$$\int_a^b f(x)dx \approx T_1 + \frac{T_2 - T_1}{3}. \tag{11}$$

Huomautus. Kaava (11) on itse asiassa ns. *Simpsonin kaava* tavanomaisesta poikkeavassa muodossa. Tavallisesti Simpsonin kaava johdetaan jakamalla väli $[a, b]$ osaväleihin, joiden lukumäärä $2n$ on parillinen, ja kirjoittamalla

$$\int_a^b f(x)dx = \sum_{k=0}^{n-1} \int_{x_{2k}}^{x_{2k+2}} f(x)dx. \tag{12}$$

Tässä kullekin integraalille $\int_{x_{2k}}^{x_{2k+2}} f(x)dx$ haetaan likiarvo siten, että approksimoidaan funktiota f välillä $[x_{2k}, x_{2k+2}]$ korkeintaan toisen asteen polynomilla, jolla on samat arvot kuin f:llä pisteissä x_{2k}, x_{2k+1} ja x_{2k+2}.

Simpsonin kaava

$$\int_a^b f(x)dx = \frac{h}{3}\big[f(x_0) + 4U + 2J + f(x_{2n})\big] + R, \tag{13}$$

missä

$$U = f(x_1) + f(x_3) + \cdots + f(x_{2n-1})$$
$$J = f(x_2) + f(x_4) + \cdots + f(x_{2n-2})$$
$$R = -\frac{(b-a)}{180} h^4 f^{(4)}(\xi)\ , \quad a \le \xi \le b$$

Näytämme, että (11) ja (13) ovat yhtäpitävät, ts.

$$T_2 + \frac{T_2 - T_1}{3} = \frac{h}{3}\left[f(x_0) + 4U + 2J + f(x_{2n})\right],$$

missä T_1 ja T_2 saadaan soveltamalla puolisuunnikassääntöä osavälien pituuksien ollessa vastaavasti n ja $2n$.

$$\begin{aligned}
T_1 &= 2h\left[\frac{f(x_0)}{2} + \qquad + f(x_2) + \qquad + f(x_4) + \cdots\right. \\
&\quad \left. + f(x_{2n-2}) + \qquad + \frac{f(x_{2n})}{2}\right] \\
4T_2 &= 4h\left[\frac{f(x_0)}{2} + f(x_1) + f(x_2) + f(x_3) + f(x_4) + \cdots\right. \\
&\quad \left. + f(x_{2n-2}) + f(x_{2n-1}) + \frac{f(x_{2n})}{2}\right] \\
4T_2 - T_1 &= h\left[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots\right. \\
&\quad \left. + 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})\right] \\
&= h\left[f(x_0) + 4U + 2J + f(x_2n)\right]
\end{aligned}$$

Huomautus 1. Tapauksessa $n = 2$ Simpsonin kaavassa $J = 0$, ja kaava voidaan kirjoittaa

$$\int_a^b f(x)dx \approx \frac{h}{3}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right].$$

Huomautus 2. Kolmannen asteen ja sitä alempaa astetta olevien polynomien integraaleille Simpsonin kaava antaa *tarkan* arvon, sillä jäännöstermi R kaavassa (13) on tällöin 0 ($f^{(4)}(x) \equiv 0$ tällaisille polynomeille).

Esimerkki 2. Sovelletaan Simpsonin kaavaa ($h = 0.1$) integraaliin

$$\int_0^1 \left(x^3 + \sin(p\sqrt{3}x)\right)dx.$$

Millä p:n arvoilla virhe on enintään $\leq \frac{1}{2} \cdot 10^{-3}$?

$$
\begin{aligned}
f(x) &= x^3 + \sin(p\sqrt{3}x) \\
f'(x) &= 3x^2 + p\sqrt{3}\cos(p\sqrt{3}x) \\
f''(x) &= 6x - 3p^2\sin(p\sqrt{3}x) \\
f'''(x) &= 6 - 3\sqrt{3}p^3\cos(p\sqrt{3}x) \\
f^{(4)}(x) &= 9p^4\sin(p\sqrt{3}x)
\end{aligned}
$$

$$
|f^{(4)}(\xi)| = |9p^4\sin(p\sqrt{3}x)| \leq 9p^4
$$

$$
\Rightarrow \quad |R| \leq \frac{b-a}{180}h^4 \max_{a\leq x\leq b}|f^{(4)}(x)| = \frac{1-0}{180}\left(\frac{1}{10}\right)^4 9p^4 = \frac{1}{20}\cdot 10^{-4}p^4.
$$

Valitaan p siten, että

$$
\begin{aligned}
\frac{1}{20}\cdot 10^{-4}p^4 &\leq \frac{1}{2}\cdot 10^{-3} \\
\Rightarrow \qquad p^4 &\leq 100 \\
\Rightarrow \qquad p^2 &\leq 10 \\
\Rightarrow \qquad |p| &\leq \sqrt{10}.
\end{aligned}
$$

Virhe on siis enintään $\frac{1}{2}\cdot 10^{-3}$ ainakin jos $|p| \leq \sqrt{10}$ edellyttäen, että laskutoimituksissa syntyviä virheitä ei oteta huomioon.

Aivan kuten kolmasosasääntöä (10) johdettaessa voidaan löytää arvio R:lle Simpsonin kaavassa edellyttäen, että funktion $f^{(4)}(x)$ arvot ovat lähellä toisiaan välillä $[a, b]$. Simpsonin kaava:

$$
\int_a^b f(x)dx = S_1 + R_1; \quad R_1 = -\frac{(b-a)}{180}h^4 f^{(4)}(\xi_1)
$$

Puolitetaan osavälit ja sovelletaan Simpsonin kaavaa uudestaan:

$$
\int_a^b f(x)dx = S_2 + R_2; \quad R_2 = -\frac{(b-a)}{180}\cdot\left(\frac{h}{2}\right)^4 f^{(4)}(\xi_2). \tag{14}
$$

Jos $f^{(4)}(\xi_1) \approx f^{(4)}(\xi_2)$, saadaan $R_1 \approx 16R_2$ ja

$$\begin{cases} \int\limits_a^b f(x)dx \approx S_1 + 16R_2 \\ \int\limits_a^b f(x)dx = S_2 + R_2 \end{cases}$$

Vähentämällä saadaan *viidestoistaosasääntö*

$$R_2 \approx \frac{S_2 - S_1}{15}. \tag{15}$$

Jos integraali halutaan laskea tietyllä tarkkuudella, voidaan käytännössä soveltaa Simpsonin sääntöä toistuvasti siten, että osavälien lukumäärä on 2, 4, 8, 16 jne. ja lopettaa laskut silloin kun kahden perättäisen tuloksen erotuksien viidestoistaosa on itseisarvoltaan pienempi kuin sallittu virhe. Lopuksi lisätään korjaus

$$\frac{S_2 - S_1}{15}$$

viimeisenä laskettaessa likiarvoa S_2.

2.2 Esimerkki

Sovelletaan puolisuunnikassääntöä ja Simpsonin sääntöä integraaliin

$$\int\limits_0^1 x^4 dx = \Big/_0^1 \frac{x^5}{5} = \frac{1}{5} = 0.2$$

Puolisuunnikassääntö yhdellä osavälillä (pituus $h = 1$):

$$T_1 = 1 \cdot \left[\frac{0^4}{2} + \frac{1^4}{2}\right] = 0.5$$

Kun osavälejä on kaksi (pituus $h = \frac{1}{2}$), saadaan

$$T_2 = \frac{1}{2}\left[\frac{0^4}{2} + \left(\frac{1}{2}\right)^4 + \frac{1^4}{2}\right] = 0.28125$$

Tässä todellinen virhe on 0.08125. Kolmasosasääntö antaa virheelle arvion

$$T_2 - \int_0^1 x^4 dx \approx -\frac{T_2 - T_1}{3} = -\frac{0.28125 - 0.50000}{3} = 0.07292$$

Kaavan (7) avulla saadaan arvio

$$|R| \leq \frac{b-a}{12} h^2 \max_{a \leq x \leq b} |f''(x)| = \frac{1-0}{12} \left(\frac{1}{2}\right)^2 \max_{0 \leq x \leq 1} |12x^2| = 0.25000$$

Kolmasosasäännön antama arvio (0.07292) todelliselle virheelle (0.08125) on aika hyvä, kun taas kaavan (7) antama arvio (0.25000) virheen ylärajalle on pessimistinen. Viidesosasäännön sekä kaavan

$$R = -\frac{b-a}{180} h^4 f^{(4)}(\xi)$$

antamassa virhearviossa on kummassakin 5 oikeaa desimaalia. Jälkimmäinen virhearvio antaa itse asiassa integraalin tarkan arvon, sillä lauseke $f^{(4)}(\xi) = 4 \cdot 3 \cdot 2 \cdot 1$ ei riipu ξ:stä.

2.3 Laskentavirheet

Laskentavirhe on se kokonaisvirhe, joka aiheutuu integroitavan funktion arvoissa $f(x)$ esiintyvistä virheistä.

Lause. *Olkoon $\tilde{f}(x_i)$ funktion f arvon $f(x_i)$ laskettu likiarvo ja ε_i siinä esiintyvä virhe, jolloin siis*

$$\tilde{f}(x_i) = f(x_i) + \varepsilon_i.$$

Oletetaan, että $|\varepsilon_i| < \varepsilon$ kaikilla i. Silloin puolisuunnikassääntöä tai Simpsonin kaavaa käytettäessä syntyvä laskentavirhe on itseisarvoltaan enintään $\varepsilon(b-a)$.

Todistus. (Puolisuunnikassäännölle) Merkitään $f_k = f(x_k)$ ja $\tilde{f}_k = \tilde{f}(x_k)$.

Laskentavirhe puolisuunnikassäännössä:

$$
\begin{aligned}
L &= h\left(\frac{\tilde{f}_0}{2} + \tilde{f}_1 + \cdots + \tilde{f}_{n-1} + \frac{\tilde{f}_n}{2}\right) - h\left(\frac{f_0}{2} + f_1 + \cdots + f_{n-1} + \frac{f_n}{2}\right) \\
&= h\left(\frac{\tilde{f}_0 - f_0}{2} + (\tilde{f}_1 - f_1) + \cdots + (\tilde{f}_{n-1} - f_{n-1}) + \frac{\tilde{f}_n - f_n}{2}\right) \\
&= h\left(\frac{\varepsilon_0}{2} + \varepsilon_1 + \cdots + \varepsilon_{n-1} + \frac{\varepsilon_n}{2}\right)
\end{aligned}
$$

Koska $|\varepsilon_i| < \varepsilon$ kaikilla i, saadaan kolmioepäyhtälön nojalla arvio

$$
\begin{aligned}
|L| &\leq h\left(\frac{|\varepsilon_0|}{2} + |\varepsilon_1| + \cdots + |\varepsilon_{n-1}| + \frac{|\varepsilon_n|}{2}\right) \\
&\leq h\left(\frac{\varepsilon}{2} + \varepsilon + \cdots + \varepsilon + \frac{\varepsilon}{2}\right) \qquad (16) \\
&\leq h\varepsilon\left(\frac{1}{2} + 1 + \cdots + 1 + \frac{1}{2}\right)
\end{aligned}
$$

Tässä esiintyvä lauseke $h\left(\frac{1}{2} + 1 + \cdots + 1 + \frac{1}{2}\right)$ saadaan tulokseksi myös silloin kun sovelletaan puolisuunnikassääntöä integraaliin $\int_a^b 1dx$. Tälle integraalille puolisuunnikassääntö antaa tarkan arvon koska jäännöstermissä esiintyvän funktion $q(x) = 1$ toinen derivaatta $q''(x) = 0$. Näin ollen

$$
\int_a^b 1dx = b - a = h\left(\frac{1}{2} + 1 + \cdots + 1 + \frac{1}{2}\right).
$$

Epäyhtälön (16) nojalla saadaan siis

$$
|L| \leq \varepsilon(b-a).
$$

Todistus on analoginen Simpsonin kaavan tapauksessa. □

2.4 Epäoleelliset integraalit

Edellä tarkasteltuja numeerisia integrointikaavoja voidaan soveltaa vain silloin kun seuraavat kaksi ehtoa on täytetty:

1. Integroimisväli $a \leq x \leq b$ on äärellinen

2. Integroitava funktio f on rajoitettu

Ongelma 1. Sovelluksissa kohdataan usein integraaleja, joissa integroimisväli on ääretön, esimerkiksi

$$\int_a^\infty f(x)dx.$$

Määritelmän mukaan

$$\int_a^\infty f(x)dx = \lim_{A\to\infty} \int_a^A f(x)dx.$$

Jos tämä raja-arvo on olemassa, niin "häntäintegraali" $\int_A^\infty f(x)dx$ saadaan miten pieneksi hyvänsä valitsemalla A kyllin suureksi.

Ongelma 1 ratkaistaan jakamalla integroimisväli äärelliseen osaväliin, jossa suoritetaan numeerinen integrointi, ja äärettömään osaväliin, jossa tehdään ns. häntäarvio.

Jos halutaan laskea $\int_a^\infty f(x)dx$ siten, että virhe on $\leq \varepsilon$, voidaan esimerkiksi valita A yhtälössä

$$\int_a^\infty f(x)dx = \int_a^A f(x)dx + \int_A^\infty f(x)dx$$

siten, että häntäintegraalin $\int_A^\infty f(x)dx$ itseisarvo on $\leq \frac{\varepsilon}{2}$. Sitten sovelletaan esimerkiksi Simpsonin kaavaa integraaliin $\int_a^A f(x)dx$ ja kaksinkertaistetaan osavälien lukumäärä niin monta kertaa, että virhe on lopulta $\frac{\varepsilon}{2}$. Tällöin saatu likiarvo integraalille $\int_a^A f(x)dx$ poikkeaa integraalista $\int_a^\infty f(x)dx$ korkeintaan ε:n verran.

Vastaavalla tavalla käsitellään integraalit

$$\int_{-\infty}^a f(x)dx \quad \text{tai} \quad \int_{-\infty}^\infty f(x)dx.$$

Esimerkki häntäarviosta

$$\int_0^\infty \frac{(\sin x)^2}{x^5+1}dx = \int_0^A \frac{(\sin x)^2}{x^5+1}dx + \int_A^\infty \frac{(\sin x)^2}{x^5+1}dx$$

Haluamme valita A:n siten, että häntäintegraali

$$\int_A^\infty \frac{(\sin x)^2}{x^5+1}dx \leq 0.001$$

Korvataan integroitava funktio majorantilla, joka on helppo integroida. Koska

$$(\sin x)^2 \leq 1 \quad \text{ja} \quad \frac{1}{x^5+1} \leq \frac{1}{x^5},$$

niin

$$\int_A^\infty \frac{(\sin x)^2}{x^5+1}dx \leq \int_A^\infty \frac{1}{x^5}dx = \Big/_{\!A}^{\infty} -\frac{1}{4x^4} = \frac{1}{4A^4}.$$

Jos valitaan $A = 4$, niin

$$\frac{1}{4A^4} = \frac{1}{4^5} = \frac{1}{1024} \leq 0.001,$$

joten

$$\int_A^\infty \frac{(\sin x)^2}{x^5+1}dx \leq 0.001$$

Huomautus. Häntäarviossa voidaan usein tyytyä arvioimaan integroitavaa funktiota melko karkeasti.

Ongelma 2. Jos integroitava funktio f ei ole rajoitettu, yritetään integraali esimerkiksi sijoituksella palauttaa jonkin rajoitetun funktion integraaliksi. Tällöin integroimisväli muuttuu usein äärettömän pitkäksi, jolloin joudutaan soveltamaan Ongelma 1:n yhteydessä kuvattua menettelyä.

Esimerkki.

$$\int_0^1 \frac{1+\sin x}{\sqrt{x}} dx$$

Tehdään sijoitus $x = \frac{1}{y}$, $dx = -\frac{1}{y^2}dy$, jolloin saadaan

$$\int_\infty^1 \frac{1+\sin\frac{1}{y}}{\sqrt{\frac{1}{y}}}\left(-\frac{dy}{y^2}\right) = \int_1^\infty \frac{1+\sin\frac{1}{y}}{y^{\frac{3}{2}}} dy$$

Näin saatu integraali on rajoitettu, kun $1 \le y < \infty$.

Toinen tapa laskea integraali on jakaa integraali osiin

$$\int_0^1 \frac{1+\sin x}{\sqrt{x}} dx = \int_0^1 \frac{dx}{\sqrt{x}} dx + \int_0^1 \frac{\sin x}{\sqrt{x}} dx$$

ja korvata $\sin x$ sen Taylorin sarjalla

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots .$$

2.5 Esimerkki

Funktion $f(x) = e^{-x^2}$ integraalifunktiota ei voida lausua alkeisfunktioiden avulla. Kuitenkin

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} = 0.88623$$

Lasketaan tämä integraali numeerisesti neljän desimaalin tarkkuudella. Kirjoitetaan

$$\int_0^\infty e^{-x^2} dx = \int_0^A e^{-x^2} dx + \int_A^\infty e^{-x^2} dx$$

ja pyritään määräämään A siten, että

$$\int_A^\infty e^{-x^2} dx \le \frac{1}{4} \cdot 10^{-4} = \frac{1}{2} \cdot \frac{1}{2} \cdot 10^{-4}.$$

Tämä voidaan tehdä eri tavoin. Jos esimerkiksi $x \geq A \geq 1$, niin

$$e^{-x^2} = \left(x^2 e^{-x^2}\right) \cdot \frac{1}{x^2} \leq A^2 e^{-A^2} \frac{1}{x^2},$$

sillä funktio $x^2 e^{-x^2}$ on vähenevä arvoilla $1 \leq x < \infty$. Siis

$$\int_A^\infty e^{-x^2} dx \leq A^2 e^{-A^2} \int_A^\infty \frac{dx}{x^2} = A^2 e^{-A^2} \cdot \frac{1}{A} = A e^{-A^2}.$$

Valitsemalla $A = 4$ saadaan

$$\int_4^\infty e^{-x^2} dx \leq 4 \cdot e^{-16} \approx 4 \cdot 10^{-7} \leq \frac{1}{4} \cdot 10^{-4}.$$

Sovelletaan puolisuunnikassääntöä integraaliin $\int_0^4 e^{-x^2} dx$.

2.6 Askelpituuden h valinnasta

Jos integroitava funktio heilahtelee voimakkaasti jollain integrointivälin $[a, b]$ osavälillä $[a, c]$, voidaan integraali kirjoittaa muotoon

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

ja laskea erikseen integraalit yli osavälien. Tällöin päästään vähemmillä laskuilla, sillä integraalia $\int_c^b f(x)dx$ laskettaessa voidaan tarvittava tarkkuus saavuttaa käyttämällä suurempaa askelpituutta h kuin välillä $[a, c]$.

2.7 Huomautus

Vaikka integraali $\int_a^b f(x)dx$ voitaisiinkin laskea analyyttisesti mikäli f on jonkun tunnetun funktion F derivaattafunktio, ts. $F' = f$, on toisinaan kannattavampaa käyttää numeerista ratkaisutapaa jolloin vältytään hankalilta laskutoimituksilta. Tällaiseen tilanteeseen päädytään esimerkiksi integraalissa

$$\int_2^4 \frac{dx}{5 - x^3}.$$

3 Yhtälöiden numeerinen ratkaiseminen

Tässä luvussa pyritään määräämään likimääräisesti yhtälön

$$f(x) = 0, \quad f \text{ jatkuva} \tag{1}$$

reaaliset juuret ja arvioimaan virhettä ko. likiarvossa.

3.1 Graafinen analyysi

Alustava kuva yhtälön (1) juurien sijainnista x-akselilla voidaan saada hahmottelemalla funktion $y = f(x)$ kuvaajan kulku xy-koordinaatistossa. Kuvaajan ja x-akselin leikkauspisteet antavat tällöin likiarvot etsityille juurille.

Toisinaan yhtälö (1) voidaan korvata yhtälöllä $f_1(x) = f_2(x)$ (mikäli esimerkiksi $f = f_1 - f_2$). Tällöin funktioiden $y = f_1(x)$ ja $y = f_2(x)$ kuvaajien leikkauspisteiden x-koordinaateista saadaan likiarvot halutuille juurille.

Esimerkki 1. Jos

$$f(x) = x^2 - \cos x,$$

niin (1) voidaan kirjoittaa muotoon

$$x^2 = \cos x.$$

Positiiviselle juurelle saadaan ensimmäinen likiarvo $x_0 = 0.8$.

3.2 Taylorin kaava

Palautetaan mieleen Taylorin kaava yksinkertaisimmissa tapauksissa.

Väliarvolause. Jos funktio f on derivoituva välillä $[a, b]$, niin ko. välillä on piste ξ siten, että

$$f(b) - f(a) = (b - a)f'(\xi). \tag{2}$$

Graafisesti tämä merkitsee sitä, että funktion $y = f(x)$ kuvaajalla on tangentti, joka on pisteiden $\big(a, f(a)\big)$ ja $\big(b, f(b)\big)$ yhdistysjanan L suuntainen. L:n kulmakerroin

$$\frac{f(b) - f(a)}{b - a}$$

on nimittäin sama kuin pisteeseen $\big(\xi, f(\xi)\big)$ piirretyn tangentin kulmakerroin $f'(\xi)$.

Seuraavaksi yksinkertaisin tapaus Taylorin kaavasta on muotoa

$$f(b) - f(a) = (b-a)f'(a) + \frac{(b-a)^2}{2} f''(\xi), \tag{3}$$

missä $a < \xi < b$. Jos merkitään $h = b - a$, niin voidaan kirjoittaa

$$f(a+h) - f(a) = hf'(a) + \frac{h^2}{2} f''(\xi). \tag{4}$$

3.3 Yleinen virhearvio

Oletetaan, että a on likiarvo yhtälön $f(x) = 0$ juurelle $\overline{x}$. Jos a on hyvä likiarvo, niin $|f(a)|$ on yleensä pieni (f jatkuva). Käänteiseen suuntaan pätee seuraava arvio.

Lause 1. *Jos $\overline{x}$ ja a ovat kuten edellä, niin*

$$|\overline{x} - a| \le \frac{|f(a)|}{\min_{x\in I} |f'(x)|},$$

missä I on väli, jonka päätepisteet ovat a ja $\overline{x}$.

Todistus. Väliarvolause (2) ja oletus $f(\overline{x}) = 0$

$$\begin{aligned}
\Rightarrow \quad & -f(a) = f(\overline{x}) - f(a) = f'(\xi)(\overline{x} - a) \\
\Rightarrow \quad & |f(a)| = |f'(\xi)||\overline{x} - a| \ge \min_{x\in I} |f'(x)||\overline{x} - a| \\
\Rightarrow \quad & |\overline{x} - a| \le \frac{|f(a)|}{\min_{x\in I} |f'(x)|}.
\end{aligned}$$

□

Huomautus. Käytännössä on otettava huomioon myös funktion arvojen laskentavirhe. Jos $\tilde{f}(a)$ on laskettu $f(a)$:n likiarvo jonka tarkkuus on $\le \varepsilon$, niin $|f(a)| \le |\tilde{f}(a)| + \varepsilon$. Silloin

$$|\overline{x} - a| \le \frac{|\tilde{f}(a)| + \varepsilon}{\min_{x\in I} |f'(x)|}.$$

Suurin mahdollinen tarkkuus saavutetaan, jos $\tilde{f}(a) = 0$. Silloin virhe on

$$|\overline{x} - a| \leq \frac{\varepsilon}{\min_{x \in I} |f'(x)|}.$$

3.4 Iterointimenetelmät

Kirjoitetaan yhtälö $f(x) = 0$ ekvivalenttiin muotoon $x = F(x)$ (kaksi yhtälöä ovat ekvivalentteja, jos niillä on samat juuret). Tämä voidaan tehdä eri tavoin, esimerkiksi

$$x = x - f(x) = F_1(x)$$

$$x = x - cf(x) = F_2(x) \qquad (c \neq 0 \text{ vakio})$$

$$x = x - g(x)f(x) = F_3(x) \quad (0 < |g(x)| < \infty)$$

Yhtälöllä $f(x) = 0$ ja $x = F_i(x)$ on samat juuret ($i = 1, 2, 3$).

Iterointimenetelmissä lähdetään liikkeelle jostakin alkuarvosta x_0 ja määritellään jono $x_1, x_2, \ldots$ *iterointikaavalla*

$$x_{n+1} = F(x_n) \qquad (n = 0, 1, 2, \ldots). \tag{5}$$

Ongelma. Milloin lukujono $\{x_n\}$ suppenee kohti etsittyä yhtälön $x = F(x)$ juurta $\overline{x}$?

Lause 2. *Jos lukujono $\{x_n\}$ suppenee kohti raja-arvoa α ja F on jatkuva, niin α on yhtälön $x = F(x)$ juuri.*

Todistus. Annetaan $n \to \infty$ yhtälössä $x_{n+1} = F(x_n)$.

Vasen puoli: $x_{n+1} \to \alpha$

Oikea puoli: Koska F on jatkuva, niin $F(x_n) \to F(\alpha)$.

Siis

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

□

Iteroinnin suppenemista koskeva peruslause on:

Lause 3. *Oletetaan, että epäyhtälö* $|F'(x)| \leq m < 1$ *on voimassa jollakin välillä, joka sisältää juuren* $\overline{x}$ *ja kaikki pisteet* x_n. *Silloin* $\lim_{n\to\infty} x_n = \overline{x}$.

Todistus.

$$\left.\begin{array}{rcl} x_1 & = & F(x_0) \\ \overline{x} & = & F(\overline{x}) \end{array}\right\} \Rightarrow \quad x_1 - \overline{x} = F(x_0) - F(\overline{x})$$

Väliarvolauseen (2) nojalla saadaan $F(x_0) - F(\overline{x}) = F'(\xi)(x_0 - \overline{x})$. Koska $|F'(\xi)| \leq m$, saadaan

$$|x_1 - \overline{x}| = |F(x_0) - F(\overline{x})| = |F'(\xi)||x_0 - \overline{x}| \leq m|x_0 - \overline{x}|.$$

Siis $|x_1 - \overline{x}| \leq m|x_0 - \overline{x}|$. Samalla tavoin saadaan edelleen

$$|x_2 - \overline{x}| = |F(x_1) - F(\overline{x})| = |F'(\xi_1)||x_1 - \overline{x}| \leq m|x_1 - \overline{x}|.$$

Koska $|x_1 - \overline{x}| \leq m|x_0 - \overline{x}|$, tästä seuraa

$$|x_2 - \overline{x}| \leq m|x_1 - \overline{x}| \leq m^2|x_0 - \overline{x}|.$$

Jatkamalla tätä menettelyä saadaan

$$|x_n - \overline{x}| \leq m|x_{n-1} - \overline{x}| \leq m^2|x_{n-2} - \overline{x}| \leq \cdots \leq m^n|x_0 - \overline{x}|.$$

Kun $n \to \infty$, niin $m^n \to 0$ (koska $m < 1$), joten oikea puoli $\to 0$. Siis $|x_n - \overline{x}| \to 0$ ja lause on todistettu. □

Esimerkki 3. Määrää yhtälön $x^3 - x - 1 = 0$ reaalinen juuri.

Esimerkiksi graafisesti nähdään, että reaalinen juuri on $\overline{x} \approx 1.3$. Kirjoitetaan yhtälö ekvivalenttiin muotoon

$$x = x^3 - 1 = F(x).$$

Funktion F derivaatta on $F'(x) = 3x^2$, joten $F'(x) > 1$ pisteen $x = 1.3$ ympäristössä. Ei siis ole odotettavissa, että lukujono $x_n = F(x_{n-1})$ suppenee.

Jos esimerkiksi $x_0 = 1.3$, niin $x_1 = 1.197$ ja $x_2 \approx 0.71506$, joten etäisyys juuresta $\overline{x}$ kasvaa. Eli $\overline{x} > 1.3$.

Esimerkki 4. Kirjoitetaan Esimerkin 3 yhtälö muotoon

$$x = F(x) = \frac{1}{x^2 - 1}.$$

Tällöin

$$F'(x) = -\frac{2x}{(x^2-1)^2},$$

joten $|F'(x)| > 1$ pisteen $x = 1.3$ ympäristössä.

Jos $x_0 = 1.3$, niin $x_1 = 1.4493$ ja $x_2 = 0.9087$. Iterointi hajaantuu.

Esimerkki 5. Kirjoitetaan Esimerkin 3 yhtälö muotoon

$$x = F(x) = (x+1)^{\frac{1}{3}}.$$

Tällöin

$$F'(x) = \frac{1}{3}(x+1)^{-\frac{2}{3}}$$

ja

$$F'(1.3) = \frac{1}{3}(2.3)^{-\frac{2}{3}} < 1.$$

Itse asiassa $|F'(x)| \leq \frac{1}{3}$ kun $x > 0$. Induktiolla nähdään, että $x_n > 0$ kaikilla $n \in \mathbb{N}$ jos $x_0 > 0$. Lauseen 3 nojalla iterointi suppenee jos $x_0 > 0$.

Jos $x_0 = 1.3$, saadaan $x_1 = 1.3200$, $x_2 = 1.3238$, $x_3 = 1.3245$, $x_4 = 1.3247$.

Nämä esimerkit osoittavat, että iteroinnin suppeneminen riippuu ratkaisevasti esityksen $x = F(x)$ valinnasta.

Lauseen 3 todistuksessa saatiin

$$|x_n - \overline{x}| \leq m^n |x_0 - \overline{x}|. \tag{6}$$

Tämän perusteella voidaan arvioida iteroinnin tarkkuutta, jos luvuille m ja $|x_0 - \overline{x}|$ tiedetään yläraja. Kuinka pystymme yleisesti arvioimaan suppeneeko iterointi jollakin alkuarvolla x_0?

Lauseessa 3 oletettiin, että ehto $|F'(x)| \leq m < 1$ on voimassa sellaisella välillä I, joka sisältää juuren $\overline{x}$ ja kaikki pisteet x_n. Miten käytännössä löydetään tällaiset m ja I?

Vastaus:

- Ensin etsitään sellainen väli $a \leq x \leq b$, joka sisältää juuren $\overline{x}$ ja lähtöarvon x_0
- I:ksi valitaan väli, jolla on sama keskipiste mutta kolme kertaa niin suuri pituus kuin välillä $a \leq x \leq b$
- m valitaan siten, että $|F'(x)| \leq m$ välillä I
- Jos tällöin $m < 1$, niin Lauseen 3 ehdot ovat voimassa välillä I, kaava (6) pätee ja iterointi suppenee
- Jos taas $m \geq 1$, yritetään uudestaan valitsemalla jokin pienempi väli $a \leq x \leq b$

Esimerkki 6. Kuinka monta iterointia on suoritettava käytettäessä kaavaa

$$x_{n+1} = \frac{-1}{x_n^2 + 2} \quad \text{ja lähtöarvoa} \quad x_0 = -0.5$$

yhtälön $x^3 + 2x + 1 = 0$ reaalisen juuren $\overline{x}$ määräämiseen neljän desimaalin tarkkuudella?

Käytetään kaavaa (6). Koska polynomi $x^3 + 2x + 1$ vaihtaa merkkiä välillä $-0.5 \leq x \leq -0.4$, niin $|x_0 - \overline{x}| \leq 0.1$ ja I voi olla esimerkiksi väli $-0.6 \leq x \leq -0.3$.

$$x = F(x) = \frac{-1}{x^2 + 2}; \quad F'(x) = \frac{2x}{(x^2 + 2)^2}$$

Kaikille $x \in I$ pätee

$$|F'(x)| \leq \frac{2 \cdot 0.6}{((-0.3)^2 + 2)^2} = \frac{2 \cdot 0.6}{2.09^2} \leq 0.3,$$

joten voidaan valita $m = 0.3$. Kaavasta (6) saadaan

$$|x_n - \overline{x}| \leq 0.3^n \cdot 0.1$$

Arvolla $n = 7$ oikea puoli on $\leq \frac{1}{2} \cdot 10^{-4}$.

Iteroinnin tarkkuutta arvioitaessa on syytä ottaa huomioon myös funktion arvojen laskemisesta aiheutuva virhe.

Lause 4. *Olkoon $\overline{x}$ yhtälön $x = F(x)$ juuri, ja olkoon ε_n virhe funktion F arvossa $F(x_n)$, so.*

$$x_{n+1} = F(x_n) + \varepsilon_n. \tag{7}$$

Oletetaan, että $|\varepsilon_n| \leq \varepsilon$ ja että

$$|F'(x)| \leq m < 1 \tag{8}$$

kaikille lukujen $\overline{x}$ ja x_n välisille arvoille x. Silloin

$$|x_{n+1} - \overline{x}| \leq \frac{m}{1-m}|x_{n+1} - x_n| + \frac{\varepsilon}{1-m} \tag{9}$$

Todistus.

$$x_{n+1} - \overline{x} \overset{(7)}{=} F(x_n) + \varepsilon_n - \overline{x} = F(x_n) + \varepsilon_n - F(\overline{x})$$

$$\Rightarrow \quad |x_{n+1} - \overline{x}| \leq |F(x_n) - F(\overline{x})| + |\varepsilon_n|$$

Väliarvolauseen ja epäyhtälön (8) nojalla $|F(x_n) - F(\overline{x})| \leq m|x_n - \overline{x}|$. Siis

$$\begin{aligned} |x_{n+1} - \overline{x}| &\leq m|x_n - \overline{x}| + |\varepsilon_n| \\ &= m|x_n - x_{n+1} + x_{n+1} - \overline{x}| + |\varepsilon_n| \\ &\leq m|x_n - x_{n+1}| + m|x_{n+1} - \overline{x}| + \varepsilon \end{aligned}$$

$$\Rightarrow \quad (1-m)|x_{n+1} - \overline{x}| \leq m|x_n - x_{n+1}| + \varepsilon$$

Koska $m < 1$, tästä seuraa (9). □

Esimerkki 7. Sovelletaan Lausetta 4 esimerkkiin 5.

$$f(x) = x^3 - x - 1; \qquad x = F(x) = (x+1)^{\frac{1}{3}}$$

$$x_0 = 1.3, \quad x_1 = 1.3200, \quad x_2 = 1.3238$$

Koska $f(1.3) = -0.103 < 0$ ja $f(1.4) = 0.344 > 0$, niin $1.3 \leq \overline{x} \leq 1.4$. Kyseisellä valillä pätee

$$|F'(x)| = \frac{1}{3(x+1)^{\frac{2}{3}}} \leq \frac{1}{3 \cdot 2.3^{\frac{2}{3}}} \leq 0.2 = m$$

Koska $x_1 = 1.3200$ sijaitsee selvästi välillä $1.3 \leq x \leq 1.4$, voidaan soveltaa Lausetta 4 arvolla $n = 1$. Jos oletetaan, että luvun x_2 laskentavirhe on enintään $\frac{1}{2} \cdot 10^{-3}$, saadaan

$$\begin{aligned} |x_2 - \overline{x}| &\leq \frac{0.2}{0.8} |1.3238 - 1.3200| + \frac{1}{0.8} \cdot \frac{1}{2} \cdot 10^{-3} \\ &\leq \frac{1}{4} \cdot 4 \cdot 10^{-3} + 0.625 \cdot 10^{-3} \leq 2 \cdot 10^{-3} \end{aligned}$$

Luvun x_2 poikkeama juuresta $\overline{x}$ on enintään 0.002.

Esimerkki 8. Oletetaan, että Lauseessa 4 $\varepsilon = 10^{-12}$ ja $m = \frac{1}{2}$. Jos halutaan määrätä $\overline{x}$ niin tarkkaan, että virhe $\leq 10^{-10}$, jatketaan iterointia kunnes

$$\frac{\frac{1}{2}}{1 - \frac{1}{2}} |x_{n+1} - x_n| + \frac{10^{-12}}{1 - \frac{1}{2}} \leq 10^{-10},$$

so. kunnes

$$|x_{n+1} - x_n| \leq 10^{-10} - 2 \cdot 10^{-12} = 0.98 \cdot 10^{-10}.$$

Huomautus. Ensimmäinen termi epäyhtälön (9) oikeanpuoleisessa lausekkeessa kuvaa katkaisuvirhettä ja toinen laskentavirhettä.

Lausetta 4 sovellettaessa valitaan jokin $\overline{x}$:n sisältävä väli I, jolla (8) pätee jollekin $m < 1$ jos $x_n \in I$, lasketaan x_{n+1} ja arvioidaan virhettä $|x_{n+1} - \overline{x}|$ kaavan(9) avulla. Jos $x_{n+1} \in I$, voidaan sama menettely toistaa jne. Iterointi lopetetaan, kun virhe on riittävän pieni.

Kaavan (9) antama yläraja x_{n+1}:n virheelle on aina vähintään yhtäsuuri kuin laskentavirhe arvossa $F(x_n)$. Todellinen x_{n+1}:n virhe voi kuitenkin olla pienempi.

3.5 Newtonin menetelmä

Kirjoitetaan yhtälö $f(x) = 0$ muotoon

$$x = x - \frac{f(x)}{f'(x)} = F(x).$$

Iterointikaava

$$\begin{cases} x_0 = \text{ lähtöarvo} \\ x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)} \; ; \quad n = 0, 1, 2, \ldots \end{cases}$$

Iterointi voidaan suorittaa graafisesti piirtämällä käyrälle $y = f(x)$ tangentti pisteeseen $\big(x_n, f(x_n)\big)$ ja määrätään ko. tangentin ja x-akselin leikkauspiste. Tangentin yhtälö on nimittäin

$$y - f(x_n) = f'(x_n)(x - x_n)$$

ja mainitun leikkauspisteen x-koordinaatti on siis

$$0 - f(x_n) = f'(x_n)(x - x_n),$$

siis

$$x = x_n - \frac{f(x_n)}{f'(x_n)},$$

so. $x = x_{n+1}$.

Esimerkki 9. $f(x) = x^3 - x - 1$

$$\begin{cases} x_0 = 1.3 \\ x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)} = x_n - \dfrac{x_n^3 - x_n - 1}{3x_n^2 - 1} = \dfrac{2x_n^3 + 1}{3x_n^2 - 1} \end{cases}$$

Laskutoimitukset ovat yksinkertaisempia kuin esimerkissä 5.

Jos $x_0 = 1.3$, saadaan $x_1 = 1.3253$, $x_2 = 1.32472$.

Esimerkki 10. Neliöjuuren laskeminen.

$\sqrt{a}$ on yhtälön $f(x) = x^2 - a = 0$ juuri. Tässä $f'(x) = 2x$, ja likiarvo juurelle saadaan jäännöslaskennalla

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right).$$

Algoritmi on käytännöllinen: jokaisessa vaiheessa on vain yksi jakolasku ja aritmeettinen keskiarvo.

Jos $a = 5$ ja $x_0 = 2$, saadaan

$$x_1 = 2.25, \quad x_2 = 2.235, \quad x_3 = 2.2361.$$

Tarkka arvo $\sqrt{5} = 2.23607$.

Esimerkki 11. Käänteisluvun määrääminen.

$\frac{1}{a}$ on yhtälön $f(x) = \frac{1}{x} - a = 0$ juuri. Tässä $f'(x) = -\frac{1}{x^2}$ ja juurelle saadaan likiarvo jäännöslaskennalla

$$x_{n+1} = x_n + \frac{\frac{1}{x_n} - a}{\frac{1}{x_n^2}} = x_n(2 - ax_n).$$

Jos $a = 13$ ja $x_0 = 0.1$, saadaan

$$x_1 = 0.07, \quad x_2 = 0.0763, \quad x_3 = 0.076918.$$

Tarkka arvo $\frac{1}{13} = 0.076923$.

Newtonin menetelmän suppenemista voidaan tarkastella Lauseen 3 avulla. Merkitään

$$F(x) = x - \frac{f(x)}{f'(x)},$$

jolloin

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

Jos $f(x)$ on kaksi kertaa derivoituva, niin iterointi suppenee siis jos

$$|F'(x)| = \left|\frac{f(x)f''(x)}{f'(x)^2}\right| \leq m < 1$$

jossakin juuren $\overline{x}$ ympäristössä ja $|x_0 - \overline{x}|$ on tarpeeksi pieni.

Lausetta 4 voidaan soveltaa iteroinnin tarkkuuden arvioimisessa. Tällöin ε on maksimaalinen laskuvirhe funktion f arvossa

$$F(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Koska x_n on kiinteä, on ko. virhe yhtäsuuri kuin virhe lausekkeessa $\frac{f(x_n)}{f'(x_n)}$.

Esimerkki 12. Sovelletaan Lausetta 4 Esimerkkiin 9.

$$\begin{aligned} f(x) = x^3 - x - 1, \quad & f'(x) = 3x^2 - 1, \qquad f''(x) = 6x \\ & f(1.3) = -0.103 < 0 \\ & f(1.4) = 0.344 > 0 \\ \Rightarrow \quad & 1.3 \leq \overline{x} \leq 1.4 \end{aligned}$$

Kuten Esimerkissä 7 arvioimme derivaattoja välillä $1.3 \leq x \leq 1.4$. Saadaan

$$|F'(x)| = \left|\frac{f(x)f''(x)}{f'(x)^2}\right| \leq \left|\frac{0.344 \cdot 6 \cdot 1.4}{4.1^2}\right| \leq 0.2$$

Jos oletetaan, että $\varepsilon = \frac{1}{2} \cdot 10^{-5}$, saadaan

$$\begin{aligned}
|x_2 - \overline{x}| &\leq \frac{0.2}{1-0.2}|1.32472 - 1.3253| + \frac{\frac{1}{2} \cdot 10^{-5}}{1-0.2} \\
&= \frac{1}{4} \cdot 0.00058 + \frac{1}{1.6} \cdot 10^{-5} \\
&= 0.000145 + 0.625 \cdot 10^{-5} \\
&\leq 0.000152.
\end{aligned}$$

Lause 5. *Oletetaan, että funktio f on kaksi kertaa derivoituva, $f(\overline{x}) = 0$ ja $f'(\overline{x}) \neq 0$. Jos x_n on kyllin lähellä juurta $\overline{x}$, niin on olemassa lukujen x_n ja $\overline{x}$ välillä sijaitseva piste ξ siten, että*

$$|x_{n+1} - \overline{x}| = \left|\frac{f''(\xi)}{2f'(x_n)}\right| \cdot |x_n - \overline{x}|^2 \tag{10}$$

Todistus. Olkoon $\Delta x = \overline{x} - x_n$ (korjaus), so. $\overline{x} = x_n + \Delta x$. Taylorin kaavasta (4) saadaan siten

$$0 = f(\overline{x}) = f(x_n + \Delta x) = f(x_n) + \Delta x f'(x_n) + \frac{(\Delta x)^2}{2} f''(\xi). \tag{11}$$

Oletuksen nojalla f' on jatkuva ja $f'(\overline{x}) \neq 0$. Jos x_n on kyllin lähellä juurta $\overline{x}$, on siis myös $f'(x_n) \neq 0$. Silloin (11) $\Rightarrow$

$$\begin{aligned}
0 &= \frac{f(x_n)}{f'(x_n)} + \Delta x + \frac{(\Delta x)^2}{2}\frac{f''(\xi)}{f'(\xi)} \\
&= x_n - x_{n+1} + \overline{x} - x_n + \frac{(\overline{x} - x_n)^2}{2}\frac{f''(\xi)}{f'(\xi)} \\
\Rightarrow \quad x_{n+1} - \overline{x} &= \frac{(\overline{x} - x_n)^2 f''(\xi)}{2f'(x_n)} \quad \Rightarrow (10).
\end{aligned}$$

□

Kaava (10) osoittaa, että Newtonin menetelmä on *kvadraattisesti suppeneva*: x_{n+1}:n virhe on verrannollinen x_n:n virheen neliöön. Jos verrannollisuuskerroin

$$\left|\frac{f''(\xi)}{2f'(x_n)}\right|$$

on suuruusluokkaa 1, niin oikeiden desimaalien lukumäärä kaksinkertaistuu jokaisessa iteraatioaskeleessa.

Yleisessä iterointikaavassa (5) suppeneminen on *lineaarista*, sillä kaavaa (10) vastaa yhtälö

$$|x_{n+1} - \overline{x}| = |F'(\xi)||x_n - \overline{x}|.$$

Tätä kutsutaan 1. kertaluvun menetelmäksi. Kvadraattisen suppenemisen tapauksessa on kyseessä 2. kertaluvun menetelmä. Yleisesti suppenemisen kertaluku määritellään suurimpana lukuna p, jolle pätee

$$\lim_{n\to\infty} \frac{|x_{n+1} - \overline{x}|}{|x_n - \overline{x}|^p} = c < \infty.$$

3.6 Sekanttimenetelmä

Jos Newtonin menetelmän iterointikaavassa esiintyvää derivaattaa $f'(x_n)$ approksimoidaan erotusosamäärällä

$$\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$

saadaan iterointikaava

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} = \frac{x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_{n-1}) - f(x_n)}. \tag{12}$$

Näin saatu *sekanttimenetelmä* eroaa tarkastellusta määritelmästä, sillä

- iterointikaava ei ole enää muotoa $x_{n+1} = F(x_n)$
- tarvitaan 2 lähtöarvoa x_0 ja x_1

Menetelmää voidaan käyttää esimerkiksi silloin, kun derivaatan $f'(x_n)$ laskeminen on vaivalloista ($f(x_n)$ voi olla määritelty implisiittisesti tms.).

Sekanttimenetelmä suppenee "superlineaarisesti", so. nopeammin kuin lineaarisesti mutta hitaammin kuin kvadraattisesti. Suppenemisen kertaluku on $p = \frac{\sqrt{5}+1}{2} \approx 1.618\ldots$

Jos kaavassa (12) pisteen $\big(x_{n-1}, f(x_{n-1})\big)$ asemasta käytetään pistettä

$$\big(x_0, f(x_0)\big),$$

saadaan ns. Regula Falsi. Tällöin kuvissa kaikki sekantit kulkevat pisteen $\big(x_0, f(x_0)\big)$ kautta. Menetelmä on huonompi kuin (12).

Toisessa menetelmän (12) muunnoksessa huolehditaan siitä, että peräkkäiset likiarvot x_n ja x_{n-1} ovat aina eri puolilla etsittyä juurta, jolloin

$$f(x_n)f(x_{n-1}) < 0.$$

Lasketaan x_{n+1} kaavan (12) avulla ja tutkitaan onko $f(x_n)f(x_{n+1}) < 0$. Jos on, suoritetaan seuraava iterointi. Jos taas $f(x_n)f(x_{n+1}) > 0$, niin seuraavassa iteroinnissa asetetaan $x_n = x_{n-1}$. Jokaisella askeleella saadaan tällöin virhearvio, sillä juuri on

- x_n:n ja x_{n-1}:n välissä
- x_{n+1}:n ja x_n:n välissä jne.

$f(x_n)$:n merkkiä arvioitaessa on kuitenkin otettava huomioon laskentavirhe. Jos $f(x_n)$:n laskentavirhe on vähintään ε, niin $f(x_n)$:n merkki on positiivinen jos lasketulle likiarvolle $\tilde{f}(x_n)$ pätee $\tilde{f}(x_n) - \varepsilon > 0$.

3.7 Hornerin kaavio

Hornerin kaavio on menetelmä, jonka avulla voidaan laskea annetun polynomin $p(x)$ ja sen derivaatan $p'(x)$ arvo annetussa pisteessä x_0.

Esimerkiksi polynomi

$$p(x) = 2x^3 + 5x^2 - 4x + 3$$

voidaan kirjoittaa muotoon

$$p(x) = \big((2x + 5)x - 4\big)x + 3. \tag{13}$$

Polynomin arvo $p(x_0)$ voidaan laskea vaiheittain seuraavasti:

$$\begin{aligned} c_1 &= 2x_0 + 5 \\ c_2 &= c_1 x_0 - 4 \\ c_3 &= c_2 x_0 + 3 \end{aligned}$$

Kaavan (13) nojalla $p(x_0) = c_3$.

Yleinen tapaus: Olkoon

$$p(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n.$$

Arvo $p(x_0)$ voidaan laskea käyttämällä palautuskaavaa:

$$\begin{cases} \quad c_0 = a_0 \\ \quad c_k = c_{k-1} x_0 + a_k \quad (1 \le k \le n) \\ p(x_0) = c_n \end{cases} \tag{14}$$

n:nen asteen polynomin arvo pisteessä x_0 voidaan siis laskea suorittamalla n kertolasku- ja n yhteenlaskutoimitusta.

Derivaatan $p'(x_0)$ laskeminen:

$$p'(x) = n a_0 x^{n-1} + (n-1) a_1 x^{n-2} + \cdots + a_{n-1}$$

$$\begin{cases} \quad d_0 = c_0 \\ \quad d_k = d_{k-1} x_0 + c_k \quad (1 \le k \le n-1) \\ p'(x_0) = d_{n-1} \end{cases} \tag{15}$$

Väite. $p(x) = (x - x_0) q(x) + c_n$, missä $q(x) = c_0 x^{n-1} + c_1 x^{n-2} + \cdots + c_{n-1}$.

Todistus.

$$\begin{aligned} (x - x_0) q(x) + c_n &= x q(x) - x_0 q(x) + c_n \\ &= c_0 x^n + c_1 x^{n-1} + \cdots + c_{n-1} x \\ &\quad - x_0 (c_0 x^{n-1} + c_1 x^{n-2} + \cdots + c_{n-1}) + c_n \\ &= c_0 x^n + (c_1 - c_0 x_0) x^{n-1} + (c_2 - c_1 x_0) x^{n-2} + \cdots \\ &\quad + (c_{n-1} - c_{n-2} x_0) x + (c_n - c_{n-1} x_0) \\ &= a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n \\ &= p(x). \end{aligned}$$

$\square$

Edellisen nojalla saadaan

$$\begin{aligned} p(x_0) &= (x_0 - x_0)q(x_0) + c_n = c_n \\ p'(x) &= (x - x_0)q'(x) + q(x) \\ \Rightarrow\ p'(x_0) &= (x_0 - x_0)q'(x_0) + q(x_0) = q(x_0) \end{aligned}$$

$p'(x_0)$ voidaan siis laskea Hornerin algoritmilla korvaamalla kertoimet a_k luvuilla $c_0, c_1, \ldots, c_{n-1}$. Edelleen

$$\begin{aligned} p''(x) &= (x - x_0)q''(x) + 2q'(x) \\ \Rightarrow\ p''(x_0) &= 2q'(x_0) \end{aligned}$$

Esimerkki 13. $p(x) = 2x^3 + 5x^2 - 4x + 3$; $x_0 = 1$.

Arvojen $p(x_0)$ ja $p'(x_0)$ laskemiseksi muodostetaan ns. Hornerin kaavio.

$$\begin{array}{rr|rrrl} & & 2 & 5 & -4 & 3 \\ c_{k-1}x_0 \rightarrow & x_0 = 1 & & 2 & 7 & 3 \\ \hline c_k \rightarrow & & 2 & 7 & 3 & 6 = p(1) \\ & & & & & \\ d_{k-1}x_0 \rightarrow & x_0 = 1 & & 2 & 9 & \\ \hline d_k \rightarrow & & 2 & 9 & 12 & = p'(1) \end{array}$$

Kaavoista (14) ja (15) nähdään, että derivaattaa $p'(x_0)$ laskettaessa luvuilla c_k on sama merkitys kuin luvuilla a_k laskettaessa arvoa $p(x_0)$. Näin ollen alimmalta riviltä voidaan lukea

$$d_0, d_1, d_2 = p'(1).$$

Esimerkki 14. Määrää kaikki yhtälön $p(x) = x^3 - 2x - 5 = 0$ reaaliset juuret.

Piirtämällä kuvaaja nähdään, että yhtälöllä on täsmälleen yksi reaalinen juuri $\overline{x}$, joka sijaitsee välillä $2 \leq x \leq 3$. Newtonin menetelmällä saadaan valitsemalla $x_0 = 2$

$$\begin{cases} x_0 = 2 \\ x_{n+1} = x_n - \dfrac{p(x_n)}{p'(x_n)} = x_n - \dfrac{x_n^3 - 2x_n - 5}{3x_n^2 - 2} \end{cases}$$

Käytetään Hornerin kaaviota arvojen $p(2)$ ja $p'(2)$ laskemiseen

$$\begin{array}{r|llll}
 & 1 & 0 & -2 & -5 \\
x_0 = 2 & & +1\cdot 2 & +2\cdot 2 & +2\cdot 2 \\
\hline
 & 1 & 2 & 2 & -1 = p(2) \\
 & & & & \\
x_0 = 2 & & +1\cdot 2 & +4\cdot 2 & \\
\hline
 & 1 & 4 & 10 & = p'(2)
\end{array}$$

Siis

$$x_1 = x_0 - \frac{p(x_0)}{p'(x_0)} = 2 - \frac{p(2)}{p'(2)} = 2 - \frac{-1}{10} = 2.1$$

Seuraavalle kierrokselle saadaan kaavio

$$\begin{array}{r|llll}
 & 1 & 0 & -2 & -5 \\
x = 2.1 & & +1\cdot 2.1 & +2.1\cdot 2.1 & +2.41\cdot 2.1 \\
\hline
 & 1 & 2.1 & 2.41 & 0.061 = p(2.1) \\
 & & & & \\
x = 2.1 & & +1\cdot 2.1 & +4.2\cdot 2.1 & \\
\hline
 & 1 & 4.2 & 11.23 & = p'(2.1)
\end{array}$$

$$x_2 = 2.1 - \frac{p(2.1)}{p'(2.1)} = 2.1 - \frac{0.061}{11.23} = 2.094568$$

Kaikki iteroinnit tämän jälkeen antavat tuloksen $x_n = 2.094551$.

Iterointia ei siis kannata jatkaa, kun huomataan, että $x_4 = x_3$.

Huomautus. Esimerkki 14 osoittaa, miten Hornerin kaaviota voidaan käyttää Newtonin menetelmän yhteydessä. Tässä erikoistapauksessa päästäisiin kuitenkin vähemmillä laskuilla kirjoittamalla iterointikaava muotoon

$$x_{n+1} = x_n - \frac{x_n^3 - 2x_n - 5}{3x_n^2 - 2} = \frac{2x^3 + 5}{3x_n^2 - 2}.$$

3.8 Iterointi useamman muuttujan tapauksessa

Esitämme Newtonin iterointikaavalle perustelun, jonka avulla menetelmä voidaan yleistää useamman muuttujan tapaukseen esimerkiksi epälineaaristen yhtälöryhmien numeeriseksi ratkaisemiseksi.

Olkoon $\overline{x}$ yhtälön $f(x) = 0$ juuri. Jos x_n on $\overline{x}$:n hyvä likiarvo, niin $\Delta x = \overline{x} - x_n$ on pieni. Taylorin kaavan (4) nojalla

$$0 = f(\overline{x}) = f(x_n + \Delta x) = f(x_n) + \Delta x f'(x_n) + \frac{(\Delta x)^2}{2} f''(\xi).$$

Jos $f''(\xi)$ ei ole kovin suuri, niin viimeinen termi on pieni, joten

$$\begin{aligned} 0 = f(\overline{x}) &\approx f(x_n) + \Delta x f'(x_n) \\ \Rightarrow \quad \Delta x &\approx -\frac{f(x_n)}{f'(x_n)} \end{aligned}$$

Siis

$$x_n + \Delta x \approx x_n - \frac{f(x_n)}{f'(x_n)} = x_{n+1},$$

joten x_{n+1} pitäisi olla parempi $\overline{x}$:n likiarvo kuin x_n. Näin saatiin analyyttinen perustelu Newtonin menetelmälle.

Kahden muuttujan tapauksessa etsitään likimääräistä ratkaisua yhtälöryhmälle

$$\begin{cases} f(x, y) = 0 \\ g(x, y) = 0 \end{cases} \tag{19}$$

Iterointimenetelmä löydetään kuten edellä käyttämällä Taylorin kaavaa.

Olkoon $(\overline{x}, \overline{y})$ yhtälöryhmän (19) tarkka ratkaisu ja (x_n, y_n) likiarvo. Silloin $\Delta x = \overline{x} - x_n$ ja $\Delta y = \overline{y} - y_n$ ovat pieniä ja $\overline{x} = x_n + \Delta x$, $\overline{y} = y_n + \Delta y$. Taylorin kaavasta saadaan

$$\begin{aligned} 0 = f(\overline{x}, \overline{y}) &= f(x_n + \Delta x, y_n + \Delta y) \\ &= f(x_n, y_n) + \Delta x D_1 f(x_n, y_n) + \Delta y D_2 f(x_n, y_n) + T_1 \\ &= f(x_n, y_n) + \Delta x \frac{\partial f}{\partial x}(x_n, y_n) + \Delta y \frac{\partial f}{\partial y}(x_n, y_n) + T_1 \end{aligned}$$

Vastaavasti

$$0 = g(\overline{x}, \overline{y}) = g(x_n, y_n) + \Delta x \frac{\partial g}{\partial x}(x_n, y_n) + \Delta y \frac{\partial g}{\partial y}(x_n, y_n) + T_2$$

Jäännöstermi T_1 on muotoa

$$
\begin{aligned}
T_1 = \frac{1}{2}\Bigg(&\Delta x^2 \frac{\partial^2}{\partial x^2} f(x_n + \xi\Delta x, y_n + \xi\Delta y) \\
&+ 2\Delta x\Delta y \frac{\partial^2}{\partial x \partial y} f(x_n + \xi\Delta x, y_n + \xi\Delta y) \\
&+ \Delta y^2 \frac{\partial^2}{\partial y^2} f(x_n + \xi\Delta x, y_n + \xi\Delta y)\Bigg) \qquad (0 < \xi < 1)
\end{aligned}
$$

T_1:n kussakin termissä on joku luvuista $(\Delta x)^2$, $(\Delta y)^2$ ja $\Delta x\Delta y$. Jos f:n toisen kertaluvun derivaatat eivät ole kovin suuria, niin T_1 (ja samoin T_2) on pieni, joten likimain

$$
\begin{cases}
0 = f(x_n, y_n) + \Delta x \dfrac{\partial f}{\partial x}(x_n, y_n) + \Delta y \dfrac{\partial f}{\partial y}(x_n, y_n) \\
0 = g(x_n, y_n) + \Delta x \dfrac{\partial g}{\partial x}(x_n, y_n) + \Delta y \dfrac{\partial g}{\partial y}(x_n, y_n)
\end{cases}
\tag{20}
$$

Ratkaistaan tästä yhtälöryhmästä Δx ja Δy, ja asetetaan

$$
x_{n+1} = x_n + \Delta x, \qquad y_{n+1} = y_n + \Delta y
$$

Tämä on Newtonin menetelmä yhtälöryhmän (19) ratkaisemiseksi.

Esimerkki 15. Tarkastellaan yhtälöryhmää

$$
\begin{cases}
x^2 - y^2 = 1 \\
x^2 + y^2 = 4
\end{cases}
$$

$$
\begin{aligned}
f(x, y) &= x^2 - y^2 - 1, \qquad & \frac{\partial f}{\partial x} &= 2x, \qquad & \frac{\partial f}{\partial y} &= -2y \\
g(x, y) &= x^2 + y^2 - 4, \qquad & \frac{\partial g}{\partial x} &= 2x, \qquad & \frac{\partial g}{\partial y} &= 2y
\end{aligned}
$$

Tällä yhtälöryhmällä on ratkaisu $\overline{x} = \sqrt{2.5} = 1.5811$ ja $\overline{y} = \sqrt{1.5} = 1.2247$. Yhtälöryhmä (20) on nyt

$$
\begin{cases}
0 = x_n^2 - y_n^2 - 1 + \Delta x \cdot 2x_n + \Delta y \cdot (-2y_n) \\
0 = x_n^2 + y_n^2 - 4 + \Delta x \cdot 2x_n + \Delta y \cdot 2y_n
\end{cases}
$$

Ratkaisu:

$$2x_n^2 - 5 + 4x_n \cdot \Delta x = 0 \qquad \Rightarrow \qquad \Delta x = \frac{5 - 2x_n^2}{4x_n}$$

$$2y_n^2 - 3 + 4y_n \cdot \Delta y = 0 \qquad \Rightarrow \qquad \Delta y = \frac{3 - 2y_n^2}{4y_n}$$

Iterointikaava:

$$\begin{cases} x_{n+1} = x_n + \Delta x = \dfrac{2x_n^2 + 5}{4x_n} = \dfrac{1}{2}x_n + \dfrac{5}{4} \cdot \dfrac{1}{x_n} \\ y_{n+1} = y_n + \Delta y = \dfrac{2y_n^2 + 3}{4y_n} = \dfrac{1}{2}y_n + \dfrac{3}{4} \cdot \dfrac{1}{y_n} \end{cases}$$

Alkuarvolla $x_0 = y_0 = 1.4$ saadaan

$$\begin{aligned} x_1 &= \frac{1}{2} \cdot 1.4 + \frac{5}{4} \cdot \frac{1}{1.4} = 1.593 \\ y_1 &= \frac{1}{2} \cdot 1.4 + \frac{3}{4} \cdot \frac{1}{1.4} = 1.236 \\ x_2 &= 1.5812 \\ y_2 &= 1.2248 \end{aligned}$$

3.9 Esimerkki

Tarkastellaan yhtälöä $p(x) = C$, missä p on polynomi. Joskus pieni häiriö p:n kertoimissa saattaa suuresti muuttaa yhtälön $p(x) = 0$ juuria.

Määritellään polynomi p esimerkiksi siten, että

$$p(x) = (x-1)(x-2)\cdots(x-20) = x^{20} - 210x^{19} + \cdots + 20!.$$

Yhtälöllä $p(x) = 0$ on tällöin juurina $x = 1, x = 2, \ldots, x = 20$.

Muutetaan polynomin p 19. asteen kerrointa -210 korvaamalla se luvulla $-210 - 2^{-23}$. Tämän tyyppisen yhtälön juurien määräämiseksi on laskuissa käytettävä huomattavasti suurempaa tarkkuutta kuin 2^{-23}.

4 Approksimointi polynomilla

4.1 Johdanto

Funktion approksimointi toisella funktiolla on tarpeen mm. seuraavista syistä:

- approksimoimalla annettua funktiota yksinkertaisemmalla funktiolla jotkin tehtävät helpottuvat (esimerkiksi funktion arvon laskeminen, numeerinen integrointi ja derivointi)

- jos funktion arvot tunnetaan alunperin vain äärellisen monessa pisteessä (esimerkiksi mittaustuloksissa), voidaan sopivasti valitun jatkuvan funktion avulla approksimoida funktion arvoja sellaisissa pisteissä, joissa niitä ei tunneta

Approksimoivina funktioina tulevat kysymykseen esimerkiksi

- polynomit

- rationaalifunktiot $\frac{P(x)}{Q(x)}$

- eksponentti- ja trigonometriset funktiot

Esimerkiksi voidaan approksimoida annettua funktiota f jollakin rationaalifunktiolla R. Jotta tällainen approksimointi antaisi välillä $a \leq x \leq b$ arvot $f(x)$ n:n desimaalin tarkkuudella, on oltava

$$\max_{a \leq x \leq b} |f(x) - R(x)| \leq \frac{1}{2} \cdot 10^{-n}.$$

Sanomme, että luku $\max_{a \leq x \leq b} |f(x) - R(x)|$ on funktioiden f ja R välinen *etäisyys välillä* $[a, b]$ tai funktion $f - R$ *normi välillä* $[a, b]$, merkitään $\|f - R\|$.

Usein käytännössä tunnetaan f:n arvot vain äärellisen monessa pisteessä $x_1, x_2, \ldots, x_n$ (esimerkiksi mittaustuloksissa). Tällöin voidaan käyttää normia

$$\|f - R\| = \max_{1 \leq i \leq n} |f(x_i) - R(x_i)|$$

ja pyrkiä siis löytämään rationaalifunktio R, jonka arvot pisteissä x_i poikkeavat kyllin vähän arvoista $f(x_i)$.

Probleema. Kun f on annettu, etsi rationaalifunktio R siten, että $\deg R \leq m$ ja

$$\|f - R\| \tag{$*$}$$

on mahdollisimman pieni.

Ratkaisu riippuu käytetystä normista $\|f - R\|$. Esimerkiksi ylläolleita maksiminormeja käytettäessä yksittäiset mittausvirheet voivat aiheuttaa suuren muutoksen ehdon $(*)$ toteuttavassa approksimaatiossa R. Siksi on usein parempi mitata funktioiden f ja R välinen etäisyys käyttämällä ns. *pienimmän neliön normia*

$$\|f - R\| = \left[\sum_{i=1}^{n} \big(f(x_i) - R(x_i) \big)^2 \right]^{\frac{1}{2}} \tag{1}$$

Huomautus. Etäisyyslausekkeessa (1) esiintyy neliöjuuri sen takia, että ns. sisätuloavaruuden aksioomat olisivat voimassa. Normi (1) toteuttaa esimerkiksi kolmioepäyhtälön

$$\|f + g\| \leq \|f\| + \|g\|,$$

missä $\|f\| = \|f - 0\|$. Vertaa normiin tasossa: jos (x_1, y_1) ja (x_2, y_2) ovat tason pisteitä, niin

$$\|(x_1, y_1) - (x_2, y_2)\| = \big[(x_1 - x_2)^2 + (y_1 - y_2)^2\big]^{\frac{1}{2}}.$$

4.2 Polynomiapproksimaatioista

Seuraavassa tarkastellaan lähinnä approksimointia polynomeilla. Tällöin on voimassa seuraava peruslause, jonka mukaan jatkossa funktioita voidaan approksimoida polynomeilla maksimoinnin suhteen mielivaltaisella tarkkuudella.

Weierstrassin approksimointilause. Jos f on välillä $[a, b]$ jatkuva funktio ja $\varepsilon > 0$, niin on olemassa polynomi $p(x)$ siten, että

$$\max_{a \leq x \leq b} |f(x) - p(x)| \leq \varepsilon.$$

Käytännön kannalta on tärkeää osata löytää sellainen approksimoiva polynomi, joka ei ole liian korkeaa astetta. Taloudellisinta on usein etsiä rationaalinen approksimaatio. Approksimaatio löydetään yleensä käyttämällä jotakin seuraavista menetelmistä:

- Taylorin kehitelmä
- interpolointi
- pienimmän neliön approksimointi
- splinit

4.3 Taylorin kehitelmä

Jos funktio f on n kertaa derivoituva pisteen α ympäristössä, niin f:llä on Taylorin kehitelmä

$$f(x) = f\big(\alpha + (x-\alpha)\big) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(x-\alpha)^k + \frac{f^{(n)}(\xi)}{n!}(x-\alpha)^n,$$

missä ξ on sillä välillä I, jonka päätepisteitä ovat x ja α. Tässä Taylorin polynomi

$$p(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(x-\alpha)^k$$

on korkeintaan astetta $n-1$ oleva polynomiapproksimaatio funktiolle f. Virhe on

$$p(x) - f(x) = -\frac{f^{(n)}(\xi)}{n!}(x-\alpha)^n$$

ja sille saadaan välillä $a \le x \le b$ yläraja

$$\|f-p\| = \max_{a\le x\le b} |f(x)-p(x)| \le \frac{\max|f^{(n)}(\xi)|}{n!} \max|x-\alpha|^n.$$

Jotta ko. yläraja olisi mahdollisimman pieni, on syytä valita $\alpha = \frac{a+b}{2}$, jolloin α on siis välin $a \le x \le b$ keskipiste.

Esimerkki 1. Haluamme approksimoida funktiota e^{-x} välillä $0 \le x \le 10$ kolmen desimaalin tarkkuudella. Taylorin kehitelmä pisteessä $\alpha = 0$:

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \\ e^{-x} &= 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots \\ \Rightarrow \quad e^{-x} &= \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} x^k + (-1)^n \frac{e^{-\xi}}{n!} x^n \end{aligned}$$

Taylorin polynomilla saadaan haluttu tarkkuus, jos

$$\max_{0\le x\le 10}\left|e^{-x}-\sum_{k=0}^{n-1}\frac{(-1)^k}{k!}x^k\right|\le \max_{0\le x\le 10}\left|(-1)^n e^{-x}\right|\max_{0\le x\le 10}\frac{|x|^n}{n!}=\frac{10^n}{n!}\le 0.5\cdot 10^{-3}$$

Pienin n:n arvo, jolla viimeinen epäyhtälö on voimassa, on $n = 32$.

Taylorin kehitelmä välin keskipisteessä $\alpha = 5$ on

$$e^{-x}=e^{-[5+(x-5)]}=e^{-5}e^{-(x-5)}=\sum_{k=0}^{n-1}\frac{(-1)^k e^{-5}}{k!}(x-5)^k+(-1)^n\frac{e^{-\xi}}{n!}(x-5)^n$$

Virheellä on yläraja

$$\max_{0\le x\le 10}\left|e^{-x}-\sum_{k=0}^{n-1}\frac{(-1)^k e^{-5}}{k!}(x-5)^k\right|\le\frac{5^n}{n!}$$

Tässä on oltava $n \ge 19$, jotta $\frac{5^n}{n!}\le\frac{1}{2}\cdot 10^{-3}$.

Taylorin kehitelmällä saadaan seuraavanlaisia polynomeja:

kehityskeskus	polynomin aste
$\alpha = 0$	31
$\alpha = 5$	18

Muilla menetelmillä voidaan ratkaista astetta 5 oleva polynomiapproksimaatio.

Huomautus. Taylorin kehitelmä ei siis yleensä anna tarkoituksenmukaista polynomiapproksimaatiota maksimoimisen suhteen. Tämä onkin luonnollista, sillä ko. menetelmä riippuu vain funktion f lokaalisesta käyttäytymisestä pisteen x ympäristössä. Toisaalta tässä *ympäristössä* kehitelmä approksimoi varsin hyvin funktiota f.

Esimerkki 2. Haluamme approksimoida funktiota $\arctan x$ välillä $-1 \le x \le 1$ kuuden desimaalin tarkkuudella. Taylorin kehitelmä pisteessä $\alpha = 0$:

$$\arctan x=\int_0^x\frac{dx}{1+x^2}=\int_0^x(1-x^2+x^4-x^6+\ldots)dx=x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+\ldots$$

Leibnizin lauseen avulla saadaan jäännöstermille arvio

$$\max_{-1\le x\le 1}\left|\arctan x-\sum_{k=0}^{n-1}\frac{(-1)^k}{2k+1}x^{2k+1}\right|\le\frac{1}{2k+1}$$

Jotta saataisiin riittävä tarkkuus $\frac{1}{2k+1}\le\frac{1}{2}\cdot 10^{-6}$, olisi käytettävä 1 999 999 asteen polynomia. Toisilla menetelmillä voidaan löytää 15. asteen polynomi, joka antaa vaaditun tarkkuuden. Käyttämällä rationaalista approksimaatiota

$$R(x)=\frac{a_0x+a_1x^3+a_3x^5}{1+a_2x^2+a_4x^4}$$

voidaan kertoimet $a_0,\ldots,a_4$ määrätä siten, että

$$\max_{|x|<1}|\arctan x-R(x)|\le\frac{1}{2}\cdot 10^{-6}.$$

4.4 Interpolointi

Olkoot $(x_0,f_0),(x_1,f_1),\ldots,(x_n,f_n)$ annettuja reaalilukupareja siten, että luvut $x_0,x_1,\ldots,x_n$ ovat keskenään erisuuret Haluamme löytää polynomin $p_n(x)$, joka saa kussakin pisteessä x_j annetun arvon f_j, so.

$$p_n(x_0)=f_0,\ p_n(x_1)=f_1,\ldots,p_n(x_n)=f_n,$$

ja jonka aste on korkeintaan n. Tällainen p_n on *interpolointipolynomi*. Pisteitä x_j kutsutaan *solmupisteiksi*. Niihin liittyvät luvut f_j voivat olla jonkin tunnetun funktion f (esimerkiksi $\ln x, \sin x$, tms.) arvoja siten, että $f(x_j)=f_j$; tällöin $p_n(x)$ on funktion f *polynomiapproksimaatio*, joka yhtyy funktioon f solmupisteissä.

Luvut f_j voivat myös olla kokeellisesti saatuja mittaustuloksia tai havaintoarvoja. Polynomia p_n käytetään hyväksi, jotta saataisiin likiarvoja funktiolle f mielivaltaisissa pisteissä x tai kokeellisesti määrättävälle funktiolle sellaisissa pisteissä x, joissa ei suoriteta mittausta. Tämä on *interpolointia* tai *ekstrapolointia* riippuen siitä, sijaitseeko mielenkiinnon kohteena oleva x:n arvo solmupisteiden välissä vai ei. Jälkimmäisessä tapauksessa menetelmä on yleensä vähemmän tarkka.

Myöhemmin nähdään, että tällainen korkeintaan astetta n oleva interpolointipolynomi on olemassa ja yksikäsitteinen. Yksikäsitteisyys seuraa siitä, että kahden mahdollisen interpolointipolynomin erotus $d_n = p_n - q_n$ on korkeintaan astetta n oleva polynomi, jolla on ainakin $n + 1$ nollakohtaa [solmupisteet $x_0, x_1, \ldots, x_n$, joissa oletuksen nojalla $p_n(x) = q_n(x)$]; siis d_n häviää identtisesti, jolloin $p_n(x) \equiv q_n(x)$.

Seuraavassa tarkastelemme eri menetelmiä $p_n(x)$:n löytämiseksi. Kaikki menetelmät antavat yksikäsitteisyyden nojalla saman polynomin, mutta polynomin esitysmuodot ja tarvittavien laskutoimitusten määrä eroavat toisistaan eri menetelmissä.

Erikoistapauksena: Lineaarinen ja kvadraattinen interpolointi

Lineaarisessa interpoloinnissa käytetään pisteiden (x_0, f_0) ja (x_1, f_1) kautta kulkevaa suoraa, johon liittyvä interpolointipolynomi on

$$p_1(x) = f_0 + (x - x_0) f[x_0, x_1], \tag{1}$$

missä

$$f[x_0, x_1] = \frac{f_1 - f_0}{x_1 - x_0} \tag{2}$$

on *ensimmäinen jaettu erotus.*

Kaavasta (1) nähdään, että $p_1(x_0) = f_0$, ja yhtälöistä (1) ja (2) seuraa $p_1(x_1) = f_1$.

Esimerkki 1. Arvioi Suomen väkilukua vuonna 1978 seuraavan taulukon avulla

Vuosi	1970	1982
Väkiluku (1000)	4598	4842

Ratkaisu.

$$p_1(1978) = 4598 + \frac{4842 - 4598}{1982 - 1970}(1978 - 1970) = 4761 \quad \text{(oik. 4758)}$$

Huomautus. $p_0(x_0) = f_0$.

Kvadraattisessa interpoloinnissa käytetään sitä korkeintaan astetta 2 olevaa polynomia $p_2(x)$, jonka kuvaaja kulkee pisteiden (x_0, f_0), (x_1, f_1) ja (x_2, f_2) kautta.

$$p_2(x) = f_0 + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2], \tag{3}$$

missä $f[x_0, x_1]$ on kuten edellä ja

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} \tag{4}$$

on *toinen jaettu erotus.*
Kaavasta (3) nähdään, että $p_2(x_0) = f_0$ ja $p_2(x_1) = f_0 + (x_1 - x_0)\frac{f_1 - f_0}{x_1 - x_0} = f_1$; lisäksi $p_2(x_2) = f_2$.

Esimerkki 2. Tiedetään, että $\ln 8.0 = 2.0794$, $\ln 9.0 = 2.1972$ ja $\ln 9.5 = 2.2513$ mutta mitä on $\ln 9.2$? Lasketaan jaetut erotukset kaavoista (2) ja (4):

$$\begin{array}{lll} x_0 = 8.0,\ f_0 = 2.0794 & & \\ & f[x_0, x_1] = 0.1178 & \\ x_1 = 9.0,\ f_1 = 2.1972 & & f[x_0, x_1, x_2] = -0.0064 \\ & f[x_1, x_2] = 0.1082 & \\ x_2 = 9.5,\ f_2 = 2.2513 & & \end{array}$$

$$\begin{aligned} (3) \Rightarrow\ p_2(x) &= 2.0794 + (x - 8.0) \cdot 0.1178 + (x - 8.0)(x - 9.0) \cdot (-0.0064) \\ &= 0.6762 + 0.2266x - 0.0064x^2 \end{aligned}$$

$p_2(9.2) = 2.2192 = \ln 9.2$ neljän desimaalin tarkkuudella.

Newtonin interpolointi ja jaetut erotukset

Kaavat (1) ja (3) ovat erikoistapauksia yleisemmästä interpolointikaavasta. Kaavan (3) perusteella p_2 saadaan p_1:stä lisäämällä siihen (3):n viimeinen termi. Samaan tilanteeseen voitaisiin pyrkiä myös yleisessä tapauksessa

$$p_n(x) = p_{n-1}(x) + g_n(x) \tag{5}$$

missä $p_{n-1}(x_0) = p_n(x_0) = f_0, \ldots, p_{n-1}(x_{n-1}) = p_n(x_{n-1}) = f_{n-1}$ ja $p_n(x_n) = f_n$. Funktion

$$g_n(x) = p_n(x) - p_{n-1}(x) \tag{5'}$$

määrittelemiseksi toteamme, että äskeisten ehtojen perusteella g_n häviää pisteissä $x_0, \ldots, x_{n-1}$. Koska g_n on korkeintaan astetta n oleva polynomi (kuten p_n), sen on siis oltava muotoa

$$g_n(x) = a_n(x - x_0)(x - x_1) \cdots (x - x_{n-1}). \tag{5”}$$

Vakion a_n määräämiseksi sijoitetaan $x = x_n$ ja ratkaistaan (5”) a_n:n suhteen. Koska (5’):n nojalla $g_n(x_n) = p_n(x_n) - p_{n-1}(x_n) = f_n - p_{n-1}(x_n)$, tulokseksi saadaan

$$a_n = \frac{f_n - p_{n-1}(x_n)}{(x_n - x_0)(x_n - x_1) \cdots (x_n - x_{n-1})} \tag{6}$$

Tapauksessa $n = 1$ on $p_{n-1}(x_1) = p_0(x_1) = f_0$, joten (6):n nojalla

$$a_1 = \frac{f_1 - p_0(x_1)}{x_1 - x_0} = \frac{f_1 - f_0}{x_1 - x_0} = f[x_0, x_1],$$

ja siten (5) antaa tulokseksi kaavan (1). Samalla tavoin arvolla $n = 2$ saadaan kaava (3) kun (6):ssa sijoitetaan

$$f_2 - p_1(x_2) \overset{(1)}{=} f_2 - f_0 - (x_2 - x_0) f[x_0, x_1],$$

jolloin

$$a_2 = \frac{f_2 - f_0 - (x_2 - x_0) f[x_0, x_1]}{(x_2 - x_0)(x_2 - x_1)} = f[x_0, x_1, x_2].$$

Kaava (3) saadaan siis kaavasta (5), kun siihen sijoitetaan arvolla $n = 2$ yhtälön (5”) mukainen $g_2(x)$. Samalla tavoin

$$a_3 = f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$$

ja yleisesti

$$a_k = f[x_0, \ldots, x_k] = \frac{f[x_1, \ldots, x_k] - f[x_0, \ldots, x_{k-1}]}{(x_k - x_0)}, \tag{7}$$

missä $f[x_0, \ldots, x_k]$ on *k:s jaettu erotus.*

Yhtälöstä (5) saadaan arvolla $n = k$

$$p_k(x) = p_{k-1}(x) + (x - x_0)(x - x_1) \cdots (x - x_{k-1}) f[x_0, \ldots, x_k]. \tag{8}$$

Antamalla k:lle peräkkäin arvot $k = 1, 2, \ldots$ tästä saadaan jälleen kaavat (1) ja (3), ja arvolla $k = n$ saadaan *Newtonin interpolointikaava*

$$\begin{aligned} f(x) \approx p_n(x) = f_0 + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2] \\ + \cdots + (x - x_0) \cdots (x - x_{n-1})f[x_0, \ldots, x_n] \end{aligned} \tag{9}$$

Todistuksen hahmotelma: Merkitään $p_{1,k}$:lla solmupisteisiin $x_1, \ldots, x_k$ kuuluvaa interpolointipolynomia (aste $\leq k$) ja $p_{0,k-1}$:lla solmupisteisiin $x_0, \ldots x_{k-1}$ kuuluvaa interpolointipolynomia (aste $\leq k$). Silloin

$$p_k(x) = \frac{(x - x_0)p_{1,k}(x) - (x - x_k)p_{0,k-1}(x)}{x_k - x_0},$$

josta saadaan p_k:hon johtavaksi kaavaksi

$$\frac{f[x_1, \ldots, x_k] - f[x_0, \ldots, x_{k-1}]}{x_k - x_0} = f[x_0, \ldots, x_k].$$

Käytännön laskuissa on hyvä käyttää seuraavan esimerkin mukaista ns. *jaettujen erotusten kaaviota.*

Esimerkki 3. Lasketaan ln 9.2:n likiarvo annettujen arvojen avulla.

x_j	$f_j = f(x_j)$	$f[x_j, x_{j+1}]$	$f[x_j, x_{j+1}, x_{j+2}]$	$f[x_j, \ldots, x_{j+3}]$
8.0	2.079442			
		0.117783		
9.0	2.197225		−0.006433	
		0.108134		0.000411
9.5	2.251292		−0.005200	
		0.097735		
11.0	2.397895			

Esimerkiksi

$$-0.005200 = \frac{0.097735 - 0.108134}{11 - 9}.$$

Kaavasta (9) saadaan

$$\begin{aligned} p_3(x) = 2.079442 + 0.117783(x - 8.0) - 0.006433(x - 8.0)(x - 9.0) \\ + 0.000411(x - 8.0)(x - 9.0)(x - 9.5), \end{aligned}$$

johon sijoittamalla $x = 9.2$ saadaan

$$f(9.2) = \ln 9.2 \approx 2.079442 + 0.141340 - 0.001544 - 0.000030 = 2.219208$$

Tarkka arvo (6 oikeaa desimaalia) on $\ln 9.2 = 2.219203$. Interpoloinnin tarkkuus kasvaa termi termiltä, sillä

$$p_1(9.2) = 2.220782; \quad p_2(9.2) = 2.219238; \quad p_3(9.2) = 2.219208.$$

Newtonin interpolointi ja eteenpäin otetut erotukset

Newtonin interpolointikaavassa (9) solmupisteiden välit ovat mielivaltaisia. Monissa sovelluksissa peräkkäisten solmupisteiden etäisyys on kuitenkin h, jolloin voidaan kirjoittaa

$$x_1 = x_0 + h, \quad x_2 = x_0 + 2h, \ \ldots \ , x_n = x_0 + nh. \tag{10}$$

Tällöin (7) ja (9) voidaan kirjoittaa yksinkertaisemmin käyttämällä ns. *eteenpäin otettuja erotuksia* seuraavasti.

Määritellään aluksi f:n ensimmäinen eteenpäin otettu erotus pisteessä x_j yhtälöllä

$$\Delta f_j = f_{j+1} - f_j,$$

toinen eteenpäin otettu erotus pisteessä x_j yhtälöllä

$$\Delta^2 f_j = \Delta f_{j+1} - \Delta f_j$$

ja yleisesti f:n k:s eteenpäin otettu erotus pisteessä x_j on

$$\Delta^k f_j = \Delta^{k-1} f_{j+1} - \Delta^{k-1} f_j \qquad (k = 1, 2, \ldots) \tag{11}$$

Jos solmupisteille x_j pätee (10), niin jaettujen erotusten ja eteenpäin otettujen erotusten välillä vallitsee yhteys

$$f[x_0, \ldots, x_k] = \frac{1}{k!h^k} \Delta^k f_0. \tag{12}$$

Todistetaan tämä induktiolla. Arvolla $k = 1$ saadaan

$$f[x_0, x_1] = \frac{f_1 - f_0}{x_1 - x_0} = \frac{1}{h}[f_1 - f_0] = \frac{1}{1!h} \Delta f_0.$$

Jos (12) pätee arvolla k, niin

$$f[x_1, \dots, x_{k+1}] = \frac{1}{k!h^k}\Delta^k f_1,$$
$$f[x_0, \dots, x_k] = \frac{1}{k!h^k}\Delta^k f_0$$

Arvolla $k+1$ saadaan siten kaavasta (7)

$$\begin{aligned} f[x_0, \dots, x_{k+1}] &= \frac{f[x_1, \dots, x_{k+1}] - f[x_0, \dots, x_k]}{x_{k+1} - x_0} \\ &= \frac{1}{x_{k+1} - x_0}\left[\frac{1}{k!h^k}\Delta^k f_1 - \frac{1}{k!h^k}\Delta^k f_0\right]. \end{aligned}$$

Tässä on (10):n nojalla $x_{k+1} - x_0 = (k+1)h$, joten

$$f[x_0, \dots, x_{k+1}] = \frac{1}{(k+1)h} \cdot \frac{1}{k!h^k}\left(\Delta^k f_1 - \Delta^k f_0\right) = \frac{1}{(k+1)!h^{k+1}}\left(\Delta^{k+1} f_0\right)$$

Siis (12) pätee myös arvolla $k+1$ m.o.t.

Asetetaan seuraavaksi kaavassa (9) $x = x_0 + rh$, missä r on positiivinen reaaliluku. Silloin $x - x_0 = rh$, $x - x_1 = (r-1)h$ (koska $x_1 - x_0 = h$) jne. Käyttämällä lisäksi kaavaa (12) saadaan Newtonin interpolointikaavalle esitys eteenpäin otettujen erotusten avulla:

$$\begin{aligned} p_n(x) &= f_0 + rh\frac{1}{1!h^1}\Delta f_0 + rh(r-1)h\frac{1}{2!h^2}\Delta^2 f_0 + \cdots \\ &\quad + rh(r-1)h\cdots(r-n+1)h\frac{1}{n!h^n}\Delta^n f_0 \\ &= f_0 + \binom{r}{1}\Delta f_0 + \binom{r}{2}\Delta^2 f_0 + \cdots + \binom{r}{n}\Delta^n f_0, \end{aligned}$$

missä *yleistetyt binomikertoimet* määritellään

$$\binom{r}{0} = 1, \quad \binom{r}{s} = \frac{r(r-1)\cdots(r-s+1)}{s!} \qquad (s > 0 \text{ kokonaisluku}) \quad (13)$$

Saadaan

$$\begin{aligned} p_n(x) &= \sum_{s=0}^{n}\binom{r}{s}\Delta^s f_0 \quad \left(x = x_0 + rh,\ r = \frac{x - x_0}{h}\right) \\ &= f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \cdots + \frac{r(r-1)\cdots(r-n+1)}{n!}\Delta^n f_0 \end{aligned} \quad (14)$$

Jos f on $n+1$ kertaa jatkuvasti derivoituva, voidaan näyttää, että kaavaan (14) liittyvä virhe on muotoa

$$\begin{aligned}\varepsilon(x) = p_n(x) - f(x) &= -\frac{1}{(n+1)!}(x-x_0)\cdots(x-x_n)f^{(n+1)}(t) \\ &= -\frac{h^{n+1}}{(n+1)!}r(r-1)\cdots(r-n)f^{(n+1)}(t),\end{aligned} \tag{15}$$

missä $f^{(n+1)}$ on f:n $(n+1)$:s derivaatta ja t on x_0:n ja x_n:n välissä (edellyttäen, että myös x sijaitsee ko. välillä).

Todistuksen idea pohjautuu Rollen lauseeseen. Määritellään apufunktio

$$Q_n(x) = f(x) - p_n(x) - \gamma(x-x_0)\cdots(x-x_n),$$

jolle pätee $Q_n(\overline{x}) = 0$. Tällöin

Q_n':lla on $n+1$ nollakohtaa
Q_n'':lla on n nollakohtaa
$\vdots$
$Q_n^{(n+1)}$:lla on 1 nollakohta t,

jolloin

$$Q_n^{(n+1)}(t) = f^{(n+1)}(t) - \gamma(n+1)! = 0 \quad \Rightarrow \quad \gamma = \frac{f^{(n+1)}(t)}{(n+1)!}$$

Esimerkki 4. Laske $\cosh 0.56$ käyttämällä kaavaa (14) ja seuraavan taulukon arvoja. Suorita virhearvio. Eteenpäin otettujen erotusten kaavio

j	x_j	$f_j = \cosh x_j$	Δf_j	$\Delta^2 f_j$	$\Delta^3 f_j$
0	0.5	1.127626			
			0.057839		
1	0.6	1.185465		0.011865	
			0.069704		0.000697
2	0.7	1.255169		0.012562	
			0.082266		
3	0.8	1.337435			

Kaavasta (14) saadaan ($x = 0.56$, joten $r = \frac{0.56-0.5}{0.1} = 0.6$)

$$\begin{aligned}\cosh 0.56 &\approx 1.127626 + 0.6 \cdot 0.057839 + \frac{0.6(-0.4)}{2} \cdot 0.011865 \\ &\quad + \frac{0.6(-0.4)(-1.4)}{6} \cdot 0.000697 \\ &= 1.127626 + 0.034703 - 0.001424 + 0.000039 \\ &= 1.160944.\end{aligned}$$

Virhearvio: Koska $\frac{d^4}{dt^4} \cosh t = \cosh t$, kaavasta (15) saadaan

$$\varepsilon_3(0.56) = -\frac{0.1^4}{4!} \cdot 0.6(-0.4)(-1.4)(-2.4) \cosh t = A \cosh t,$$

missä $A = 0.00000336$ ja $0.5 \leq t \leq 0.8$. Koska $\cosh t$ on välillä $[0.5, 0.8]$ kasvava funktio, tästä seuraa

$$A \cosh 0.5 \leq \varepsilon_3(0.56) \leq A \cosh 0.8.$$

Koska $f(x) = p_3 - \varepsilon_3(x)$, saadaan edelleen

$$p_3(0.56) - A \cosh 0.8 \leq f(x) = \cosh 0.56 \leq p_3(0.56) - A \cosh 0.5.$$

Lasketaan ylä- ja alarajojen numeroarvot:

$$1.160939 \leq \cosh 0.56 \leq 1.160941$$

Tarkka arvo (6 desimaalia) on $\cosh 0.56 = 1.160941$.

Lagrangen interpolointikaava

Oletetaan, että interpolointitehtävään liittyvissä reaalilukupareissa (x_0, f_0), $\ldots, (x_n, f_n)$ solmupisteiden x_j välit ovat mielivaltaisia. Määritellään polynomit $l_0(x), l_1(x), \ldots, l_n(x)$ yhtälöllä

$$\begin{aligned}l_0(x) &= (x - x_1)(x - x_2) \cdots (x - x_n) \\ l_k(x) &= (x - x_0) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n) \\ l_n(x) &= (x - x_0)(x - x_1) \cdots (x - x_{n-1})\end{aligned}$$

Näiden avulla muodostettu ns. *Lagrangen interpolointikaava* on

$$f(x) \approx L_n(x) = \sum_{k=0}^{n} \frac{l_k(x)}{l_k(x_k)} f_k \tag{19}$$

Oikeanpuoleisessa summassa kukin termi edustaa polynomia, joka saa yhdessä solmupisteessä arvon f_k ja muissa solmupisteissä arvon 0. Näin ollen $L_n(x_k) = f_k$ kaikille k, joten L_n on annettuihin lukupareihin liittyvä interpolointipolynomi.

Esimerkki 7. Etsitään $\ln 9.2$ soveltamalla Lagrangen interpolointikaavaa (19) ja seuraavan taulukon arvoja

x	9.0	9.5	10.0	11.0
$\ln x$	2.19722	2.25129	2.30259	2.39790

$$(19) \Rightarrow \quad L_3(x) = \frac{l_0(x)}{l_0(x_0)}f_0 + \frac{l_1(x)}{l_1(x_1)}f_1 + \frac{l_2(x)}{l_2(x_2)}f_2 + \frac{l_3(x)}{l_3(x_3)}f_3,$$

missä

$$\begin{aligned} l_0(x) &= (x-9.5)(x-10)(x-11) \\ l_1(x) &= (x-9)(x-10)(x-11) \text{ jne.} \end{aligned}$$

Siten

$$\begin{aligned} \ln 9.2 \approx{}& \frac{-0.43200}{-1.00000} \cdot 2.19722 + \frac{0.28800}{0.37500} \cdot 2.25129 \\ &+ \frac{0.10800}{-0.50000} \cdot 2.30259 + \frac{0.04800}{3.00000} \cdot 2.39790 \\ ={}& 2.21920 \quad \text{(5 oikeaa desimaalia)} \end{aligned}$$

Käytännön laskuissa polynomien L_n käyttö on hankalaa, sillä laskutoimitukset ovat työläitä ja polynomien astetta kasvatettaessa aikaisempi laskutyö menee hukkaan. Lagrangen polynomeilla on kuitenkin huomattava teoreettinen merkitys.

4.5 Pienimmän neliön approximointi

Esimerkki 1. Annettu 5 tason pistettä (x_i, y_i) seuraavasti:

i	x_i	y_i
1	1.0	1.0
2	1.5	1.7
3	2.0	2.2
4	2.5	2.5
5	3.0	2.5

Etsimme toisen asteen polynomia $p_2(x) = a + bx + cx^2$ siten, että

$$\sum_{i=1}^{5} \big(y_i - p_2(x_i)\big)^2$$

on mahdollisimman pieni. Toisin sanoen, haluamme löytää vakiot a, b ja c siten, että

$$F(a,b,c) = \sum_{i=1}^{5} \big(y_i - a - bx_i - cx_i^2\big)^2$$

on mahdollisimman pieni. Välttämätön ääriarvoehto

$$\frac{\partial F}{\partial a} = \frac{\partial F}{\partial b} = \frac{\partial F}{\partial c} = 0$$

johtaa lineaariseen yhtälöryhmään, ns. *normaaliyhtälöihin*

$$\begin{cases} -2\displaystyle\sum_{i=1}^{5} (y_i - a - bx_i - cx_i^2) = 0 \\ -2\displaystyle\sum_{i=1}^{5} x_i(y_i - a - bx_i - cx_i^2) = 0 \\ -2\displaystyle\sum_{i=1}^{5} x_i^2(y_i - a - bx_i - cx_i^2) = 0 \end{cases}$$

Matriisimuodossa:

$$\begin{pmatrix} \sum 1 & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix}$$

missä kaikki summausindeksit juoksevat 1:stä 5:teen. Kerroinmatriisi on symmetrinen ja voidaan osoittaa, että yhtälöryhmällä on yksikäsitteinen ratkaisu, joka antaa funktiolle $F(a,b,c)$ pienimmän arvon. Kerroinmatriisi ja oikeanpuoleinen vektori lasketaan käyttämällä taulukkoa:

x_i	y_i	x_i^2
1	1.0	1.0
1.5	1.7	2.25
2.0	2.2	4.0
2.5	2.5	6.25
3	2.5	9.0

Muodostetaan ensin summat

$$\sum_{i=1}^{5} x_i = 10, \quad \sum_{i=1}^{5} y_i = 9.9, \quad \sum_{i=1}^{5} x_i^2 = 22.5$$

Lasketaan sitten taulun pystyvektorien "sisätulot"

$$\sum_{i=1}^{5} x_i^3 = 55, \quad \sum_{i=1}^{5} x_i^4 = 142.125, \quad \sum_{i=1}^{5} x_i y_i = 21.7, \quad \sum_{i=1}^{5} x_i^2 y_i = 51.75$$

Koska $\sum_{i=1}^{5} 1 = 5$, voidaan normaaliyhtälöt kirjoittaa

$$\begin{pmatrix} 5 & 10 & 22.5 \\ 10 & 22.5 & 55 \\ 22.5 & 55 & 142.125 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9.9 \\ 21.7 \\ 51.75 \end{pmatrix}$$

Ryhmä ratkaistaan esimerkiksi Gaussin eliminointikeinolla. Ratkaisu on

$$a = -1.1400, \quad b = 2.5886, \quad c = -0.4571.$$

Esimerkki 2. Sama probleema kuin Esimerkissä 1, mutta $p_2(x)$ kirjoitetaan muotoon

$$p_2(x) = a' + b'(x-2) + c'(x-2)^2.$$

Tällöin saadaan yhtälöryhmä

$$\begin{pmatrix} \sum 1 & 0 & \sum (x_i-2)^2 \\ 0 & \sum (x_i-2)^2 & 0 \\ \sum (x_i-2)^2 & 0 & \sum (x_i-2)^4 \end{pmatrix} \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum (x_i-2) y_i \\ \sum (x_i-2)^2 y_i \end{pmatrix}$$

jonka kertoimista muutamat ovat nollia symmetrian vuoksi. Keskimmäisestä yhtälöstä saadaan b', kun taas yhtälöt 1 ja 3 muodostavat yhtälöparin, jossa tuntemattomina ovat a' ja c'.

Huomautus. Yritteessä kannattaa yleensäkin käyttää sellaisia muuttujan $x - \alpha$ polynomeja missä α sijaitsee arvojen x_i keskivaiheilla.

Esimerkit 1 ja 2 kuvaavat ns. *pienimmän neliösumman* approksimointia (PNS): funktiota $y = f(x)$ approksimoidaan polynomilla $p_2(x)$ siten, että virheen $-(f(x) - p_2(x))$ pisteissä $x_1, x_2, \ldots, x_5$ saamien arvojen neliösumma on mahdollisimman pieni.

Esimerkissä 1 olisi lausekkeen $F(a,b,c)$ asemasta voitu yrittää minimoida esimerkiksi funktiota

$$F_1(a,b,c) = \sum_{i=1}^{5} |y_i - a - bx_i - cx_i^2|$$

tai "maksiminormia" (vrt. §4.1)

$$F_2(a,b,c) = \max_{1\leq i\leq 5} |y_i - a - bx_i - cx_i^2|.$$

Tällaiset approksimointitehtävät ovat laskennallisesti hankalampia kuin pienimmän neliösumman keino.

Huomautus. Esimerkissä 1 ei ole olemassa toisen asteen polynomia, joka saa pisteessä x_i arvon y_i kaikille i. Tässä ei siis etsitä interpolointipolynomia.

Jos halutaan approksimoida annettua funktiota f n:nnen asteen polynomilla, missä $n \neq 2$, voidaan menetellä kuten Esimerkeissä 1 ja 2. Tällöin normaaliyhtälöitä on $n+1$ kappaletta, sillä n:nnen asteen polynomilla on $n+1$ kerrointa. Tapauksessa $n = 1$ approksimoivan polynomin kuvaaja on suora, ns. *pienimmän neliösumman suora.*

Jos funktio f tunnetaan koko välillä $[a,b]$ ja halutaan f:lle pienimmän neliön approksimointi n:nnen asteen polynomilla, voidaan yrittää minimoida

$$F(c_0, c_1, \ldots, c_n) = \int_a^b \Big(f(x) - \sum_{k=0}^{n} c_k x^k\Big)^2 dx.$$

Tässä siis mitataan "virhettä" integraalilla summan asemasta.

Normaaliyhtälöiden ratkaiseminen saattaa olla numeerisesti hankalaa siksi, että pienet häiriöt kertoimissa voivat suuresti muuttaa ko. yhtälöiden ratkaisuja. Sen vuoksi on usein syytä käyttää muita menetelmiä approksimaation löytämiseksi erityisesti mikäli n on suuri.

Eräs keino on käyttää yritteenä polynomin $\sum_{k=0}^{n} c_k x^k$ asemasta lineaarikombinaatiota

$$p(x) = \sum_{k=0}^{n} a_k L_k(x),$$

missä astetta k olevat polynomit $L_k(x)$ on valittu siten, että normaaliyhtälöiden kerroinmatriisi muuttuu diagonaalimatriisiksi. Tällaiset polynomit ovat annettuun approksimointitehtävään liittyviä *ortogonaalisia polynomeja.*

Havainnollistamme yllä kuvattua menettelyä esimerkin avulla. Yksinkertaisuuden vuoksi käytämme "integraalinormia"

$$\|f-p\| = \Big(\int_{-1}^{1} \big(f(x)-p(x)\big)^2 dx \Big)^{\frac{1}{2}},$$

mutta samaan tapaan voidaan tarkastella "summanormia"

$$\|f-p\| = \Big(\sum_{i=1}^{m} \big(f(x_i)-p(x_i)\big)^2 \Big)^{\frac{1}{2}}.$$

Esimerkki 3. Approksimoidaan funktiota $\sin \pi x$ kolmannen asteen polynomilla välillä $-1 \le x \le 1$ siten, että

$$\int_{-1}^{1} \left(\sin \pi x - \sum_{k=0}^{3} c_k x^k \right)^2 dx$$

on mahdollisimman pieni. Tehdään yrite

$$p(x) = \sum_{k=0}^{3} a_k L_k(x),$$

missä

$$L_0(x) = 1, \quad L_1(x) = x, \quad L_2(x) = \frac{3}{2}x^2 - \frac{1}{2} \quad \text{ja} \quad L_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x.$$

Lausekkeen

$$F(a_0, a_1, a_2, a_3) = \int_{-1}^{1} \left(\sin \pi x - \sum_{k=0}^{3} a_k L_k(x) \right)^2 dx$$

minimoimiseksi asetetaan

$$\frac{\partial F}{\partial a_k} = 0 \qquad (k = 0, 1, 2, 3).$$

Tällöin saadaan normaaliyhtälöt

$$\begin{cases} 2\int\limits_{-1}^{1} L_0\Big(\sin \pi x - \sum\limits_{k=0}^{3} a_k L_k(x)\Big)dx = 0 \\ 2\int\limits_{-1}^{1} L_1\Big(\sin \pi x - \sum\limits_{k=0}^{3} a_k L_k(x)\Big)dx = 0 \\ 2\int\limits_{-1}^{1} L_2\Big(\sin \pi x - \sum\limits_{k=0}^{3} a_k L_k(x)\Big)dx = 0 \\ 2\int\limits_{-1}^{1} L_3\Big(\sin \pi x - \sum\limits_{k=0}^{3} a_k L_k(x)\Big)dx = 0 \end{cases} \tag{1}$$

Kertoimien a_0, a_1, a_2, a_3 ratkaisemiseksi (1) voidaan kirjoittaa myös muotoon

$$\begin{cases} a_0\int\limits_{-1}^{1} L_0^2 dx + a_1\int\limits_{-1}^{1} L_0 L_1 dx + a_2\int\limits_{-1}^{1} L_0 L_2 dx + a_3\int\limits_{-1}^{1} L_0 L_3 dx = \int\limits_{-1}^{1} L_0 \sin \pi x dx \\ a_0\int\limits_{-1}^{1} L_1 L_0 dx + a_1\int\limits_{-1}^{1} L_1^2 dx + a_2\int\limits_{-1}^{1} L_1 L_2 dx + a_3\int\limits_{-1}^{1} L_1 L_3 dx = \int\limits_{-1}^{1} L_1 \sin \pi x dx \\ a_0\int\limits_{-1}^{1} L_2 L_0 dx + a_1\int\limits_{-1}^{1} L_2 L_1 dx + a_2\int\limits_{-1}^{1} L_2^2 dx + a_3\int\limits_{-1}^{1} L_2 L_3 dx = \int\limits_{-1}^{1} L_2 \sin \pi x dx \\ a_0\int\limits_{-1}^{1} L_3 L_0 dx + a_1\int\limits_{-1}^{1} L_3 L_1 dx + a_2\int\limits_{-1}^{1} L_3 L_2 dx + a_3\int\limits_{-1}^{1} L_3^2 dx = \int\limits_{-1}^{1} L_3^2 \sin \pi x dx \end{cases} \tag{2}$$

Yhtälöryhmän (2) kerroinmatriisi on diagonaalimatriisi, sillä

$$\int\limits_{-1}^{1} L_i L_j dx = 0, \quad \text{kun } i \neq j,$$

ja

$$\int_{-1}^{1} L_i^2 dx = \frac{1}{i+\frac{1}{2}} \quad (0 \le i \le 3).$$

Ratkaisu on näin ollen

$$a_k = (k+\frac{1}{2}) \int_{-1}^{1} L_k(x) \sin \pi x dx \qquad (0 \le k \le 3)$$

Esimerkiksi

$$a_0 = \frac{1}{2} \int_{-1}^{1} \sin \pi x dx = 0; \quad a_1 = \frac{3}{2} \int_{-1}^{1} x \sin \pi x dx = \frac{3}{\pi} \qquad \text{jne.}$$

Huomautus. Polynomit $L_k(x)$ ovat ns. *Legendren polynomeja.* Ne voidaan määritellä palautuskaavalla

$$L_{k+1}(x) = \frac{2k+1}{k+1} x L_k(x) - \frac{k}{k+1} L_{k-1}(x); \quad L_0(x) = 1, \quad L_1(x) = x.$$

4.6 Splinit

Interpolointipolynomin asteluvun korottaminen ei aina paranna interpoloinnin tarkkuutta. Esimerkiksi funktion

$$f(x) = \frac{1}{1+25x^2}$$

välillä $[-1, 1]$ muodostettujen interpolointipolynomien solmupisteet voidaan valita siten, että interpoloinnin maksimivirhe kasvaa rajatta polynomin asteluvun kasvaessa. Tällainen epästabiilisuus voidaan välttää käyttämällä ns. *splinejä.*

Splinit ovat ns. *palapolynomeja*, so. jatkuvia funktioita joiden rajoittuma kahden peräkkäisen solmupisteen määräämälle välille on polynomi. Lisäksi vaaditaan, että solmupisteissä funktio on riittävän monta kertaa derivoituva. Interpolointi tällaisilla funktioilla on yleensä numeerisesti stabiilia.

Yksinkertaisimman esimerkin approksimoinnista palapolynomeilla antavat paloittain lineaariset funktiot (esimerkiksi puolisuunnikassääntö). Tällaisen funktion kuvaaja on kuitenkin särmikäs, eikä funktio ole derivoituva särmien kohdalla. Splineihin päädytään vaatimalla derivoituvuus myös solmupisteissä.

Käytännössä tärkeimmät lienevät ns. *kuutiosplinit*, jotka määritellään seuraavasti. Olkoon f funktio, jota halutaan approksimoida välillä $a \leq x \leq b$. Oletetaan, että väli on jaettu osaväleihin, joiden päätepisteinä ovat seuraavat *solmupisteet*:

$$a = x_0 < x_1 < \cdots < x_n = b. \tag{1}$$

Solmupisteisiin (1) liittyvä kuutiosplini on välillä $a \leq x \leq b$ kaksi kertaa jatkutvasti derivoituva funktio, jonka rajoittuma kullekin kahden peräkkäisen solmupisteen rajoittamalle välille on korkeintaan astetta 3 oleva polynomi. Funktiota f *interpoloiva kuutiosplini* g saadaan vaatimalla lisäksi, että

$$g(x_0) = f(x_0) = f_0,\ g(x_1) = f(x_1) = f_1, \ldots, g(x_n) = f(x_n) = f_n. \tag{2}$$

Seuraavassa osoitamme, että tällainen ehdot (2) täyttävä kuutiosplini on aina olemassa. Se ei kuitenkaan ole yksikäsitteinen, sillä derivaatoille välin $a \leq x \leq b$ päätepisteissä voidaan antaa mielivaltaiset arvot:

Lause 1. *Olkoon f välillä $a \leq x \leq b$ määritelty funktio. Oletetaan, että välillä $a \leq x \leq b$ on annettu solmupisteet (1), ja olkoot k_0 ja k_n mielivaltaisia reaalilukuja. Silloin on olemassa täsmälleen yksi solmupisteisiin (1) liittyvä kuutiosplini g siten, että (2) pätee ja lisäksi*

$$g'(x_0) = k_0, \quad g'(x_n) = k_n. \tag{3}$$

Todistus. Helposti nähdään, että jokaisella osavälillä $I_j = [x_j, x_{j+1}]$ löytyy täsmälleen yksi korkeintaan astetta 3 oleva polynomi $p_j(x)$ siten, että

$$p_j(x_j) = f(x_j), \quad p_j(x_{j+1}) = f(x_{j+1}) \tag{4}$$

ja p_j' saa osavälin päätepisteissä ennalta annetut arvot

$$p_j'(x_j) = k_j, \quad p_j'(x_{j+1}) = k_{j+1}. \tag{5}$$

Tällaisen polynomin lauseke saadaan merkitsemällä $c_j = \frac{1}{x_{j+1}-x_j}$ seuraavaan muotoon

$$
\begin{aligned}
p_j(x) = {} & f(x_j)c_j^2(x-x_{j+1})^2[1+2c_j(x-x_j)] \\
& + f(x_{j+1})c_j^2(x-x_j)^2[1-2c_j(x-x_{j+1})] \\
& + k_jc_j^2(x-x_j)(x-x_{j+1})^2 + k_{j+1}c_j^2(x-x_j)^2(x-x_{j+1})
\end{aligned} \tag{6}
$$

Yksikäsitteisyys on harjoitustehtävä.

Etsittävän kuutiosplinin $g(x)$ rajoittuma välille I_j on muotoa (6). Jotta $g(x)$ olisi kaksi kertaa jatkuvasti derivoituva myös solmupisteissä x_j, on oltava

$$p''_{j-1}(x_j) = p''_j(x_j) \qquad (1 \le j \le n-1) \tag{7}$$

Kaavasta (6) saadaan derivoimalla

$$
\begin{aligned}
p''_j(x) = {} & 2f(x_j)c_j^2[1+4c_j(x-x_{j+1})+2c_j(x-x_j)] \\
& + 2f(x_{j+1})c_j^2[1-4c_j(x-x_j)-2c_j(x-x_{j+1})] \\
& + (4k_j+2k_{j+1})c_j^2(x-x_{j+1}) + (4k_{j+1}+2k_j)c_j^2(x-x_j)
\end{aligned} \tag{8}
$$

Yhdistämällä kaava (8) ja ehdot (7) saadaan

$$
\begin{aligned}
& 6f(x_{j-1})c_{j-1}^2 - 6f(x_j)c_{j-1}^2 + (4k_j+2k_{j-1})c_{j-1} \\
& = -6f(x_j)c_j^2 + 6f(x_{j+1})c_j^2 - (4k_j+2k_{j+1})c_j \quad (i \le j \le n-1)
\end{aligned} \tag{9}
$$

Merkitsemällä $\Delta f_{j-1} = f(x_j) - f(x_{j-1})$ ja $\Delta f_j = f(x_{j+1}) - f(x_j)$ kuten luvussa § 4.4 kaava (9) voidaan kirjoittaa muodossa

$$c_{j-1}k_{j-1} + 2(c_{j-1}+c_j)k_j + c_jk_{j+1} = 3(c_{j-1}^2\Delta f_{j-1} + c_j^2\Delta f_j) \tag{10}$$

Yhtälöt (10) muodostavat arvoilla $1 \le j \le n-1$ lineaarisen yhtälöryhmän, jossa tuntemattomia ovat luvut $k_1, \dots, k_{n-1}$. Tämän yhtälöryhmän kerroinmatriisi on säännöllinen (ilman todistusta), joten systeemillä (10) on täsmälleen yksi ratkaisu $k_1, \dots, k_{n-1}$. Sijoittamalla tämä ratkaisu yhtälöihin (6) saadaan polynomit $p_0, p_1, \dots, p_{n-1}$, joiden välillä $a \le x \le b$ määrittelemä funktio $g(x)$ on ainoa ehdot (3) ja (7) toteuttava funktiota $f(x)$ interpoloiva kuutiosplini. $\square$

Analysoimalla Lauseen 1 todistusta saadaan algoritmi splinin määrittelemiseksi. Yksinkertaisuuden vuoksi tarkastellaan tapausta, jossa peräkkäisten solmupisteiden etäisyys on vakio h, so.

$$x_1 = x_0 + h, \quad x_2 = x_0 + 2h, \ \ldots, \ x_n = nh$$

Silloin $c_j = \frac{1}{x_{j+1}-x_j} = \frac{1}{h}$, joten kertomalla (10) puolittain h:lla ja merkitsemällä $f(x_j) = f_j$ saadaan

$$k_{j-1} + 4k_j + k_{j+1} = \frac{3}{h}(f_{j+1} - f_{j-1}) \quad (i \leq j \leq n-1) \tag{11}$$

Tässä k_0 ja k_n ovat annettuja, esimerkiksi $k_0 = f'(a)$, $k_n = f'(b)$. Algoritmin ensimmäisessä vaiheessa ratkaistaan $k_1, \ldots, k_{n-1}$ lineaarisesta yhtälöryhmästä (11). Toisessa vaiheessa määritellään splinin $g(x)$ kertoimet. Välillä $x_j \leq x \leq x_{j+1} = x_j + h$ splini $g(x)$ voidaan kehittää Taylorin polynomiksi

$$p_j(x) = a_{j0} + a_{j1}(x - x_j) + a_{j2}(x - x_j)^2 + a_{j3}(x - x_j)^3, \tag{12}$$

missä

$$\begin{cases} a_{j0} = p_j(x_j) = f_j \\ a_{j1} = p'_j(x_j) = k_j \qquad \text{(5):n nojalla} \\ a_{j2} = \dfrac{1}{2}p''_j(x_j) = \dfrac{3}{h^2}(f_{j+1} - f_j) - \dfrac{1}{h}(k_{j+1} + 2k_j) \end{cases} \tag{13}$$

(vertaa kaavan (9) oikea puoli). Kertoimien a_{j3} laskemiseksi todetaan, että (8):n perusteella

$$p''_j(x_{j+1}) = \frac{6}{h^2}\,(f_j - f_{j+1}) + \frac{2}{h}(2k_{j+1} + k_j)$$

kun taas (12):n nojalla

$$p''_j(x_{j+1}) = 2a_{j2} + 6a_{j3}h$$

Merkitsemällä oikeat puolet yhtäsuuriksi saadaan yhtälö, josta a_{j3} voidaan ratkaista:

$$a_{j3} = \frac{1}{3h}\left\{\frac{3}{h^2}(f_j - f_{j+1}) + \frac{1}{h}(k_j + 2k_{j+1}) - a_{j2}\right\}$$

Kun tähän sijoitetaan a_{j2}:n lauseke kaavasta (13), saadaan lopulta

$$a_{j3} = \frac{2}{h^3}(f_j - f_{j+1}) + \frac{1}{h^2}(k_{j+1} + k_j). \tag{14}$$

Esimerkki 1. Etsitään funktiota $f(x) = x^4$ välillä $-1 \leq x \leq 1$ interpoloiva kuutiosplini $g(x)$, kun solmupisteet ovat $x_0 = -1$, $x_1 = 0$, $x_2 = 1$ ja välin päätepisteissä vaaditaan, että $g'(-1) = f'(-1)$, $g'(1) = f'(1)$.

Koska $n = 2$, yhtälöryhmä (11) surkastuu yhtälöksi

$$k_0 + 4k_1 + k_2 = \frac{3}{h}(f_2 - f_0).$$

Tässä $k_0 = f'(-1) = -4$, $k_2 = f'(1) = 4$, $h = 1$, $f_2 = f(1) = 1$ ja $f_0 = f(-1) = 1$, joten yhtälön ratkaisuksi saadaan $k_1 = 0$. Kaavoista (13) ja (14) seuraa siten arvolla $j = 0$

$$\begin{aligned}
a_{00} &= f_0 = 1\\
a_{01} &= k_0 = -4\\
a_{02} &= 3(f_1 - f_0) - (k_1 + 2k_0) = 3(0-1) - (0-8) = 5\\
a_{03} &= 2(f_0 - f_1) + (k_1 + k_0) = 2(1-0) + (0-4) = -2
\end{aligned}$$

Näin ollen

$$p_0(x) = 1 - 4(x+1) + 5(x+1)^2 - 2(x+1)^3 = -x^2 - 2x^3.$$

Samalla tavoin saadaan arvolla $j = 1$

$$\begin{aligned}
a_{10} &= f_1 = 0\\
a_{11} &= k_1 = 0\\
a_{12} &= 3(1-0) - (4 + 2\cdot 0) = -1\\
a_{13} &= 2(0-1) + 4 + 0 = 2
\end{aligned}$$

Siis

$$p_2(x) = -x^2 + 2x^3.$$

Etsitty splini $g(x)$ toteuttaa siis ehdot

$$g(x) = \begin{cases} -x^2 - 2x^3, & \text{kun } -1 \leq x \leq 0\\ -x^2 + 2x^3, & \text{kun } \quad 0 \leq x \leq 1 \end{cases}$$

Esimerkki 2. Funktiosta f tiedetään seuraavat arvot

$$f_0 = f(0) = 1, \quad f_1 = f(2) = 9, \quad f_2 = f(4) = 41, \quad f_3 = f(6) = 41.$$

Etsitään interpoloiva kuutiosplini, joka toteuttaa ehdot $k_0 = 0$, $k_3 = -12$.

Koska $n = 3$ ja $h = 2$, yhtälöryhmä (11) on

$$\begin{aligned} k_0 + 4k_1 + k_2 &= \frac{3}{2}(f_2 - f_0) = 60 \\ k_1 + 4k_2 + k_3 &= \frac{3}{2}(f_3 - f_1) = 48 \end{aligned}$$

Arvoilla $k_0 = 0$, $k_3 = -12$ ratkaisuksi saadaan $k_1 = 12$, $k_2 = 12$. Kaavoista (13) ja (14) saadaan splinin kertoimiksi arvolla $j = 0$

$$\begin{aligned} a_{00} &= f_0 = 1 \\ a_{01} &= k_0 = 0 \\ a_{02} &= \frac{3}{4}(f_1 - f_0) - \frac{1}{2}(k_1 + 2k_0) = \frac{3}{4}(9 - 1) - \frac{1}{2}(12 + 0) = 0 \\ a_{03} &= \frac{2}{8}(f_0 - f_1) + \frac{1}{4}(k_1 + k_0) = \frac{2}{8}(1 - 9) + \frac{1}{4}(12 + 0) = 1 \end{aligned}$$

Välillä $0 \le x \le 2$ pätee siis

$$g(x) = p_0(x) = 1 + x^3.$$

Arvolla $j = 1$

$$\begin{aligned} g(x) = p_1(x) &= 9 + 12(x-2) + 6(x-2)^2 - 2(x-2)^3 \\ &= 25 - 36x + 18x^2 - 2x^3 \qquad (2 \le x \le 4) \end{aligned}$$

Arvolla $j = 2$

$$\begin{aligned} g(x) = p_2(x) &= 41 + 12(x-4) - 6(x-4)^2 \\ &= -103 + 60x - 6x^2 \qquad (4 \le x \le 6) \end{aligned}$$

Splineillä on seuraava minimiominaisuus:

Lause 2. *Olkoon f välillä $a \le x \le b$ kaksi kertaa jatkuvasti derivoituva funktio, ja olkoon g solmupisteisiin (1) liittyvä funktiota f interpoloiva kuutiosplini siten, että*

$$g'(a) = f'(a) \quad \text{ja} \quad g'(b) = f'(b). \tag{15}$$

Silloin

$$\int_a^b f''(x)^2 dx \geq \int_a^b g''(x)^2 dx, \tag{16}$$

ja yhtäsuuruus pätee vain jos $f(x) \equiv g(x)$ *välillä* $a \leq x \leq b$.

Todistus. Yhtälöstä (15) seuraa osittaisintegroinnilla

$$\begin{aligned}\int_a^b g''(x)\,[f''(x) - g''(x)]\,dx &= \sum_{j=0}^{n-1} \int_{x_j}^{x_{j+1}} p_j''(x)\,[f''(x) - g''(x)]\,dx \\ &= \sum_{j=0}^{n-1} \left\{ \Big/_{x_j}^{x_{j+1}} p_j''(x)\,[f'(x) - g'(x)] - \int_{x_j}^{x_{j+1}} p_j'''(x)\,[f'(x) - g'(x)]\,dx \right\}\end{aligned}$$

Tässä viimeinen integraali häviää, koska $p_j'''(x)$ on vakio ja

$$\int_{x_j}^{x_{j+1}} [f'(x) - g'(x)]\,dx = \Big/_{x_j}^{x_{j+1}} [f(x) - g(x)] = 0.$$

Myös sijoitustermien summa on nolla, sillä

$$\begin{aligned}&\sum_{j=0}^{n-1} \Big/_{x_j}^{x_{j+1}} p_j'(x)\,[f'(x) - g'(x)] \\ &= p_0'(x_1)[f'(x_1) - g(x_1)] - p_0'(x_0)[f'(x_0) - g'(x_0)] \\ &\quad + p_1'(x_2)[f'(x_2) - g(x_2)] - p_1'(x_1)[f'(x_1) - g'(x_1)] + \cdots = 0\end{aligned}$$

kaavan (15) nojalla. Siis

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

josta seuraa

$$\int_a^b [f''(x) - g''(x)]^2 dx = \int_a^b f''(x)^2 dx - 2\int_a^b f''(x)g''(x)dx + \int_a^b g''(x)^2 dx$$

$$= \int f''(x)^2 dx - \int_a^b g''(x)^2 dx$$

Koska vasen puoli on ei-negatiivinen, nähdään että (15) pätee ja että yhtäsuuruus on voimassa vain jos $f''(x) - q''(x) \equiv 0$, jolloin $f(x) \equiv g(x)$ välillä $[a, b]$. $\square$

4.7 Polynomiapproksimaatioiden käyttömahdollisuuksia

1° **Funktion approksimointi tietokoneessa**

Esimerkiksi funktio $\sin x$.

2° **Numeerinen integrointi**

Laskettaessa likiarvoa integraalille $\int_a^b f(x)dx$ voidaan $f(x)$ korvata jollakin approksimoivalla polynomilla $p(x)$ siten, että virhe

$$\left| \int_a^b \big(f(x) - p(x)\big) dx \right|$$

on kyllin pieni. Jos käytetään pisteisiin $x_k = a + k\frac{b-a}{n}$ $(0 \le k \le n)$, liittyvää interpolointipolynomia, jolloin siis $p(x_k) = f(x_k)$, saadaan ns. n:nnen kertaluvun Newtonin-Cotesin integrointikaava. Arvolla $n = 2$ saadaan Simpsonin kaava.

Toisinaan on paikallaan käyttää katkaistua Taylorin kehitelmää. Esimerkiksi käyttämällä esitystä

$$\frac{\sin x}{x} \approx 1 - \frac{x^2}{3!} + \frac{x^4}{5!}$$

saadaan laskettua integraalin

$$\int_0^{\frac{1}{2}} \frac{\sin x}{x} dx$$

arvo tarkkuudella $\pm 10^{-6}$.

3° Numeerinen derivointi

Jos $p(x)$ is funktiota $f(x)$ approksimoiva polynomi, niin funktion derivaattaa $f'(x)$ voidaan approksimoida polynomilla $p'(x)$. Jos halutaan approksimaatio kokonaisella välillä $a \leq x \leq b$, niin $p(x)$ voi olla esimerkiksi interpoloiva kuutiosplini tai pienimmän neliön approksimointi. Interpolointipolynomeja käytettäessä voi virhe $|f'(x) - p'(x)|$ kasvaa rajatta polynomin asteluvun kasvaessa.

Lokaalinen likiarvo derivaatalle $f'(x)$ löydetään derivoimalla sopivaa interpolointipolynomia. Kun funktiota approksimoidaan pisteitä $\big(x-h, f(x-h)\big)$ ja $\big(x+h, f(x+h)\big)$ yhdistävällä suoralla, saadaan likiarvo

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}.$$

Tässä esiintyvä virhe voidaan laskea käyttämällä Taylorin kaavaa. Vähentämällä puolittain yhtälöt

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h}{2}f''(x) + \frac{h^3}{6}f'''(\xi_1) \\ f(x-h) &= f(x) - hf'(x) + \frac{h}{2}f''(x) - \frac{h^3}{6}f'''(\xi_2) \end{aligned}$$

saadaan

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{6}f'''(\xi), \tag{1}$$

sillä

$$\frac{f'''(\xi_1) + f'''(\xi_2)}{2} = f'''(\xi)$$

jollakin ξ (mikäli $f'''(x)$ on jatkuva).

Derivaatan $f''(x)$ approksimointiin voidaan käyttää toisen asteen interpolointipolynomia, jonka kuvaaja kulkee pisteiden

$$\big(x-h, f(x-h)\big), \quad \big(x, f(x)\big) \quad \text{ja} \quad \big(x+h, f(x+h)\big)$$

kautta. Kaava (1) vastaava kaava on tällöin

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{h^2}{12} f^{(4)}(\xi). \tag{2}$$

Kaavoissa (1) ja (2) esiintyvät jäännöstermit saadaan miten pieniksi hyvänsä pienentämällä h:ta. Näin ollen derivaatoille $f'(x)$ ja $f''(x)$ voidaan saada (periaatteessa) mielivaltaisen tarkat approksimaatiot käyttämällä funktion arvoja pisteissä $f(x)$ ja $f(x \pm h)$, missä h on kyllin pieni. Jos arvot $f(x+h)$ ja $f(x-h)$ voidaan laskea tarkkuudella $\pm\varepsilon$ ja $|f'''(x)| \leq M$ välillä $[x-h, x+h]$, niin kokonaisvirhe derivaatan $f'(x)$ approksimaatiossa

$$\frac{f(x+h) - f(x-h)}{2h}$$

on kaavan (1) nojalla enintään

$$\frac{\varepsilon + \varepsilon}{2h} + \frac{h^2}{6} M = \frac{\varepsilon}{h} + \frac{h^2 M}{6}.$$

Katkaisuvirheestä $-\frac{h^2}{6} f'''(\xi)$ aiheutuva termi $\frac{h^2}{6} M$ pienenee kun $h \to 0$, kun taas funktion arvojen epätarkkuudesta aiheutuva termi $\frac{\varepsilon}{h}$ kasvaa kun $h \to 0$. Kokonaisvirheen minimoimiseksi olisi h valittava niin, että

$$T(h) = \frac{\varepsilon}{h} + \frac{h^2 M}{6}$$

on mahdollisimman pieni. Asettamalla $T'(h) = 0$ saadaan

$$h = \left(\frac{3\varepsilon}{M} \right)^{\frac{1}{3}}.$$

Toinen keino on asettaa molemmat virhetermit yhtäsuuriksi:

$$\frac{\varepsilon}{h} = \frac{h^2 M}{6} \quad \Rightarrow \quad h = \left(\frac{6\varepsilon}{M} \right)^{\frac{1}{3}}.$$

5 Differentiaaliyhtälöt

5.1 Johdanto

Esitämme joitakin menetelmiä alkuarvotehtävän

$$\begin{cases} y' = f(x, y) \\ y(a) = \eta \end{cases} \tag{1}$$

ratkaisemiseksi. Ratkaisulta $y(x)$ vaaditaan, että se on derivoituva jollakin annetulla välillä $a \leq x \leq b$ ja saa pisteessä $x = a$ arvon η. Lisäksi $y'(x) = f\big(x, y(x)\big)$.

Lause 1. *(Picardin lause) Oletetaan, että funktio f on jatkuva ja että*

$$\left| \frac{\partial f(x, y)}{\partial y} \right| \leq K$$

kaikilla $x \in [a, b]$ ja kaikilla y. Silloin alkuarvotehtävällä (1) on yksikäsitteinen ratkaisu.

Esimerkki 1. Osoitetaan, että alkuarvotehtävällä

$$\begin{cases} y' = \dfrac{x^2 \sin y}{1 + x^2} \\ y(0) = 1 \end{cases}$$

on yksikäsitteinen ratkaisu välillä $0 \leq x < \infty$.

Funktio $f(x, y) = \frac{x^2 \sin y}{1+x^2}$ on kaikkialla differentioituva ja siten myös jatkuva.

$$\left| \frac{\partial f}{\partial y} \right| = \left| \frac{x^2 \cos y}{1 + x^2} \right| \leq \frac{x^2}{1 + x^2} \leq 1.$$

Lauseen 1 ehto on siis täytetty.

Seuraavassa tarkastellaan ns. askelmenetelmiä, joissa lasketaan likiarvoja ratkaisulle $y(x)$ äärellisen monessa välin $a \leq x \leq b$ pisteessä $x_1, x_2, \ldots, x_N$. Ratkaisu voidaan tietenkin löytää myös suljetussa muodossa. Tämä on kuitenkin harvinaista ja saattaa johtaa niin hankaliin laskuihin, että numeerinen ratkaisu on edullisempi.

Esimerkki 2. Alkuarvotehtävällä

$$\begin{cases} y' = x + \dfrac{2y}{1-x^4} \\ y(0) = 1 \end{cases}$$

on ratkaisu

$$y(x) = \left(\frac{1+x}{1-x}\right)^{\frac{1}{2}} e^{\arctan x} \left(\int_0^x u \left(\frac{1-u}{1+u}\right)^{\frac{1}{2}} e^{-\arctan u} du + 1 \right).$$

5.2 Askelmenetelmät

Askelmenetelmistä yksinkertaisin on *Eulerin menetelmä*. Tarkastellaan alkuarvotehtävää

$$\begin{cases} y' = f(x,y) \\ y(a) = \eta \end{cases} \tag{1}$$

välillä $a \le x \le b$. Jaetaan väli N:ään osaväliin, joiden pituus on $h = \frac{b-a}{N}$ ja päätepisteet

$$x_k = a + kh \qquad (0 \le k \le N).$$

Merkitään y_k:lla ratkaisun $y(x)$ likiarvoa pisteessä $x = x_k$. Virhe pisteessä x_k on siis $y_k - y(x_k)$. Eulerin menetelmässä lasketaan likiarvot $y_1, y_2, \ldots, y_N$ peräkkäin käyttämällä kaavaa

$$y_{n+1} = y_n + hf(x_n, y_n). \tag{2}$$

Tällöin $x_0 = a$ ja $y_0 = y(a) = \eta$, joten

$$\begin{aligned} y_1 &= y + hf(a, \eta) \\ y_2 &= y_1 + hf(x_1, y_1) \\ y_3 &= y_2 + hf(x_2, y_2) \\ &\vdots \\ y_N &= y_{N-1} + hf(x_{N-1}, y_{N-1}) \end{aligned} \tag{3}$$

Esimerkki 3. Käsitellään Esimerkin 2 alkuarvotehtävää Eulerin menetelmällä. Käyttämällä askelpituutta $h = 0.1$ saadaan

$$
\begin{aligned}
y(0) &= y_0 = 1 \\
y(0.1) &\approx y_1 = 1 + 0.1 \cdot 2 = 1.2 \\
y(0.2) &\approx y_2 = 1.2 + 0.1 \left(0.1 + \frac{2 \cdot 1.2}{1 - 10^{-4}} \right) = 1.45 \\
y(0.3) &\approx y_3 = 1.45 + 0.1 \left(0.2 + \frac{2 \cdot 1.45}{1 - 16 \cdot 10^{-4}} \right) = 1.76 \\
y(0.4) &\approx y_4 = 2.118 \\
y(0.5) &\approx y_5 = 2.593
\end{aligned}
$$

Perustellaan Eulerin menetelmää ensin analyyttisesti ja sitten graafisesti. Likiarvo $y_0 = \eta$ saadaan alkuehdosta $y(a) = \eta$. Oletetaan, että ratkaisu $y(x)$ on kaksi kertaa jatkuvasti derivoituva. Taylorin kaavastaa saadaan

$$
y(x_0 + h) = y(x_0) + hy'(x_0) + \frac{h^2}{2} y''(\xi).
$$

Jättämällä tässä pois jäännöstermi saadaan arvolla $y(x_0 + h) = y(x_1)$ likiarvo

$$
y_1 = y(x_0) + hy'(x_0) = \eta + hf(x_0, y_0) = \eta + f(a, \eta).
$$

Näin saatiin ensimmäinen kaavoista (3). Virhe likiarvossa y_1

$$
y_1 - y(x_1) = -\frac{h^2}{2} y''(\xi)
$$

Koska y'' on jatkuva, tämä virhe saadaan miten pieneksi hyvänsä valitsemalla h tarpeeksi pieneksi. Toinen kaavoista (3) perustellaan samalla tavoin, ja yleisesti kaavasta

$$
y(x_n + h) = y(x_n) + hy'(x_n) + \frac{h^2}{2} y''(\xi_n)
$$

saadaan likiarvo

$$
y_{n+1} = y_n + hf(x_n, y_n)
$$

jättämällä pois $\frac{h^2}{2} y''(\xi_n)$ ja korvaamalla $y'(x_n) = f\big(x_n, y(x_n)\big)$ likiarvolla $f(x_n, y_n)$. Jälkimmäisestä toimenpiteestä aiheutuvaa virhettä voidaan arvioida väliarvolauseen avulla, jos Lauseen 1 ehto

$$
\left| \frac{\partial f(x, y)}{\partial y} \right| \leq K
$$

on täytetty:

$$\left|f\big(x_n, y(x_n)\big) - f(x_n, y_n)\right| \le K\,|y(x_n) - y_n|$$

Jos lisäksi $|y''(x)| \le M$ kaikilla $x \in [a, b]$, voidaan osoittaa, että globaalille katkaisuvirheelle pätee

$$|y(x_n) - y_n| \le \frac{hM}{2}\frac{e^{K(x_n-a)} - 1}{K}.$$

Virhe oikeanpuoleisessa päätepisteessä $x_N = b$ on siis enintään

$$\frac{hM}{2K}\left(e^{K(b-a)} - 1\right),$$

jos laskentavirhettä ei oteta huomioon.

Eulerin menetelmä

Probleema:

$$\begin{cases} y' = f(x, y) \\ y(a) = \eta \end{cases}$$

Algoritmi:

$$\begin{aligned} y_{n+1} &= y_n + hf(x_n, y_n) \\ y_0 &= \eta \\ x_n &= a + nh, \quad h = \frac{b-a}{N} \end{aligned} \tag{4}$$

Virhearvio:

$$|y(x_n) - y_n| \le \frac{hM}{2K}\left(e^{K(x_n-a)} - 1\right).$$

Jos Lauseen 1 ehdot on täytetty, differentiaaliyhtälöllä $y' = f(x, y)$ on täsmälleen yksi annetun pisteen (x, y) kautta kulkeva integraalikäyrä. Eulerin menetelmässä piirretään integraalikäyrälle tangentti pisteeseen (x_n, y_n), jolloin y_{n+1} on ko. tangentin ja suoran $x = x_{n+1}$ leikkauspisteen y-koordinaatti:

$$y_{k+1} = y_k + hf(x_n, y_n)$$

Huomautus. Eulerin menetelmän sijasta käytetään tavallisesti tarkempia menetelmiä, joissa virheen ylärajalla on samantapainen lauseke, mutta h:n sijasta esiintyy tekijä h^p, missä $p > 1$. Suuremman tarkkuuden ansiosta voidaan tällöin käyttää suurempaa askelpituutta h, jolloin tarvittavien askeleiden lukumäärä $N = \frac{b-a}{N}$ vähenee.

Eulerin menetelmä on esimerkki *yksiaskelmenetelmästä*. Sellaista menetelmää, jossa likiarvon y_{n+1} laskemiseen käytetään useampia aikaisempia likiarvoja, esimerkiksi y_n, y_{n-1}, y_{n-2}, $\ldots, y_{n-p}$, kutsutaan *moniaskelmenetelmiksi*. Tällainen on esimerkiksi *kaksiaskelmenetelmä*

$$y_{n+1} = y_{n-1} + 2hf(x_n, y_n). \tag{5}$$

Perustelu: Integroidaan yhtälö $y'(x) = f\big(x, y(x)\big)$ yli välin $[x_{n-1}, x_{n+1}]$ ja approksimoidaan integraalia

$$\int\limits_{x_{n-1}}^{x_{n+1}} f\big(t, y(t)\big)dt$$

tulolla

$$f\big(x_n, y(x_n)\big)\big(x_{n+1} - x_{n-1}\big) = 2hf\big(x_n, y(x_n)\big),$$

jolloin saadaan

$$y(x_{n+1}) - y(x_{n-1}) = \int\limits_{x_{n-1}}^{x_{n+1}} f\big(t, y(t)\big)dt \approx 2hf\big(x_n, y(x_n)\big)$$

Jos tässä tarkat arvot $y(x_k)$ korvataan likiarvolla y_k, saadaan (5). Voidaan näyttää, että n:n askeleen jälkeen virheelle pätee

$$|y(x_n) - y_n| \leq M \cdot h^2,$$

missä M on vakio.

Huomautus. Moniaskelmenetelmät vaativat erillisen käynnistysmenetelmän, sillä alussa tiedetään vain yksi alkuehto $y_0 = \eta$. Esimerkiksi kaavaa (5) käytettäessä tarvitaan arvon y_2 laskemiseen y_0:n lisäksi myös y_1, sillä $y_2 = y_0 + 2hf(x_1, y_1)$. Käynnistysmenetelmänä voidaan käyttää esimerkiksi jotakin yksiaskelmenetelmää.

Heunin menetelmä eli parannettu Eulerin menetelmä

Olkoon

$$y_{n+1}^{(e)} = y_n + hf(x_n, y_n)$$

Eulerin menetelmän antama likiarvo pisteessä x_{n+1} ja

$$k_1 = hf(x_n, y_n)$$

vastaava funktion arvon muutos. k_1:lle saadaan toinen likiarvo approksimoimalla integraalikäyrää pisteeseen (x_n, y_n) piirretyn tangentin asemasta sekantilla, jonka suunta on integraalikäyrän tangentin suunta pisteessä $(x_{n+1}, y_{n+1}^{(e)})$:

$$k_2 = hf(x_{n+1}, y_{n+1}^{(e)}) = hf(x_n + h, y_n + k_1)$$

Heunin menetelmä:

$$y_{n+1} = y_n + \frac{1}{2}(k_1 + k_2)$$

Katkaisuvirhe Heunin menetelmässä on luokkaa $O(h^2)$, ts. virhe $\leq M \cdot h^2$.

Klassinen Runge-Kutta-menetelmä:

$$\begin{aligned}
k_1 &= hf\big(x_n, y_n\big) \\
k_2 &= hf\big(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\big) \\
k_3 &= hf\big(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\big) \\
k_4 &= hf\big(x_n + h, y_n + k_3\big)
\end{aligned}$$

$$y_{n+1} = y_n + \frac{1}{6}\big(k_1 + 2k_2 + 2k_3 + k_4\big)$$

Katkaisuvirhe on suuruusluokkaa $O(h^4)$.

Esimerkki. Lasketaan alkuarvotehtävän

$$\begin{cases} y' = xy \\ y(0) = 1 \end{cases}$$

ratkaisukäyrän $y(x)$ likiarvo pisteessä $x = 0.4$ käyttämällä klassista Runge-Kutta-menetelmää ja askelpituutta $h = 0.4$.

Nyt $f(x,y) = xy$, joten saadaan

$$\begin{aligned}
k_1 &= hx_0y_0 = 0 \\
k_2 &= h\big(x_0 + \frac{h}{2}\big)\big(y_0 + \frac{k_1}{2}\big) = h \cdot \frac{h}{2} \cdot 1 = 0.08 \\
k_3 &= h\big(x_0 + \frac{h}{2}\big)\big(y_0 + \frac{k_2}{2}\big) = h \cdot \frac{h}{2}(1 + 0.04) = 0.0832 \\
k_4 &= h\big(x_0 + h\big)\big(y_0 + k_3\big) = 0.173312
\end{aligned}$$

$$y(0.4) \approx y_1 = 1 + \frac{1}{6}\big(0 + 2 \cdot 0.08 + 2 \cdot 0.0832 + 0.173312\big) = 1.083285$$

Katkaisuvirhe $\approx 2 \cdot 10^{-6}$.

5.3 Implisiittiset menetelmät

Integroidaan yhtälön

$$y'(x) = f\big(x, y(x)\big)$$

vasen puoli yli välin $[x_n, x_{n+1}]$, jolloin saadaan

$$\int_{x_n}^{x_{n+1}} y'(t)dt = y(x_{n+1}) - y(x_n).$$

Arvioidaan oikean puolen integraalia

$$\int_{x_n}^{x_{n+1}} f\big(x, y(x)\big)dx$$

likimääräisesti puolisuunnikassäännöllä:

$$\int_{x_n}^{x_{n+1}} f\big(x, y(x)\big)dx \approx \frac{h}{2}\left[f\big(x_n, y(x_n)\big) + f\big(x_{n+1}, y(x_{n+1})\big)\right]$$

Korvaamalla vasemmalla ja oikealla tarkat arvot $y(x_n)$ ja $y(x_{n+1})$ likiarvoilla y_n ja y_{n+1} saadaan:

Trapetsikaava

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

Globaalinen katkaisuvirhe on $O(h^2)$.

Trapetsikaava on esimerkki *implisiittisestä* menetelmästä: y_{n+1} ei riipu eksplisiittisesti muun muassa luvuista x_n, y_n kuten tähän asti tarkastelluissa *eksplisiittisissä* menetelmissä, vaan y_{n+1}:n laskemiseksi on ratkaistava yhtälö (1), joka saattaa olla epälineaarinen.

Esimerkki. Tarkastellaan vielä probleemaa $y' = xy$, $y(0) = 1$. Differentiaaliyhtälön lineaarisuuden takia trapetsimenetelmässä ratkaistava yhtälö (1) on lineaarinen:

$$y_{n+1} = y_n + \frac{h}{2}\big[x_n y_n + x_{n+1} y_{n+1}\big].$$

Käyttämällä askelpituutta $h = 0.2$ saadaan

$$\begin{aligned} & y_1 = y_0 + 0.1\big[x_n y_n + x_{n+1} y_{n+1}\big] = 1 + 0.1\big(0 + 0.2 y_1\big) \\ \Leftrightarrow\ & y_1 = 1 + 0.02 y_1 \\ \Leftrightarrow\ & y_1 \approx 1.0204 \end{aligned}$$

Arvolla $n = 1$ saadaan vastaavasti

$$\begin{aligned} & y_2 = y_1 + 0.1[x_1 y_1 + x_2 y_2] = 1.0204 + 0.1[0.2 \cdot 1.0204 + 0.4 \cdot y_2] \\ \Rightarrow\ & y(0.4) \approx y_2 = \frac{1.0408}{1 - 0.04} \approx 1.0842 \end{aligned}$$

Katkaisuvirhe on $\approx 9 \cdot 10^{-4}$.

Esimerkki. Tarkastellaan alkuarvotehtävää

$$\begin{cases} y' = e^{-y} \\ y(0) = 1 \end{cases}$$

Differentiaaliyhtälö on nyt epälineaarinen. Arvolla $n = 0$ kaava (1) on

$$y_1 = y_0 + \frac{h}{2}\big[e^{-y_0} + e^{-y_1}\big] = 1 + \frac{h}{2}\big[e^{-1} + e^{-y_1}\big].$$

Jos esimerkiksi $h = 0.2$, niin y_1 on ratkaistava yhtälöstä

$$y_1 = 1 + 0.1[e^{-1} + e^{-y_1}].$$

Määritellään apufunktio $g(y)$ siten, että

$$g(y_1) = y_1 - 0.1e^{-y_1} - (1 + 0.1e^{-1}) = 0,$$

jolloin ratkaisu löytyy esimerkiksi Newtonin menetelmällä. Lähtöarvoksi käy Eulerin menetelmän antama likiarvo

$$y_1^{(o)} = y_0 + he^{-y_0} \approx 1.0736.$$

Newtonin menetelmällä saadaan tällöin

$$\begin{aligned} y_1^{(1)} &= 1.071053 \\ y_1^{(2)} &= 1.071053 \end{aligned}$$

Siis $y_1 = 1.071053$, ja voidaan siirtyä tapaukseen $n = 1$.

Trapetsikaavaa käytettäessä y_{n+1} ratkaistaan muotoa $g(y) = 0$ olevasta yhtälöstä, missä

$$g(y) = y - \frac{h}{2} f(x_{n+1}, y) - \left[y_n + \frac{h}{2} f(x_n, y_n) \right].$$

Jos ratkaisu tapahtuu Newtonin menetelmällä kuten yllä olevassa esimerkissä, joudutaan kullakin iteraatioaskeleella laskemaan derivaatta

$$g'(y) = 1 - \frac{h}{2} \frac{\partial f}{\partial y}(x_{n+1}, y).$$

Vaihtoehtoinen tapa yhtälön $g(y) = 0$ ratkaisemiseksi on kiintopisteiterointi:

$$\begin{aligned} y_{n+1}^{(0)} &= y_n + hf(x_n, y_n) \\ y_{n+1}^{(k+1)} &= y_n + \frac{h}{2} \left[f(x_n, y_n) + f(x_{n+1}, y_{n+1}^{(k)}) \right] \quad (k = 0, 1, 2, \ldots) \end{aligned} \tag{2}$$

Lähtöarvo on tässäkin laskettu Eulerin menetelmällä, joka toimii ns. ennustajana. Saatua likiarvoa korjataan trapetsikaavalla. Algoritmi on esimerkki ns. *ennustaja-korjaaja-menetelmästä*. § 3.4:n mukaan riittävä ehto iteroinnin (2) suppenemiselle on epäyhtälön

$$\left| \frac{h}{2} \frac{\partial f}{\partial y} \right| < 1 \tag{3}$$

voimassaolo jossakin pisteen (x_{n+1}, y_{n+1}) ympäristössä. Askelpituuden h on siis oltava kyllin pieni.

Esimerkki. Tarkastellaan alkuarvotehtävän

$$\begin{cases} y' = e^{-y} \\ y(0) = 1 \end{cases}$$

ratkaisemista ennustaja-korjaaja-menetelmällä (2). Iterointi suppenee kohti yhtälön (1) ratkaisua, jos

$$\left|\frac{h}{2}\frac{\partial f}{\partial y}\right| = \frac{h}{2}e^{-y} < 1$$

jossakin pisteen (x_{n+1}, y_{n+1}) ympäristössä. Koska $e^{-y} > 0$ kaikilla y:n arvoilla, kaavoissa (2) kaikki termit ovat positiivisia. Koska lisäksi $y(0) = y_0 = 1$, nähdään induktiolla, että $y_{n+1}^{(k+1)} \geq 1$ kaikille k ja n. Ehto (3) on siis voimassa pisteen (x_{n+1}, y_{n+1}) ympäristössä, jos esimerkiksi $h = 0.1$.

Jos $\left|\frac{\partial f}{\partial y}\right|$ on suuri, on iteroinnoin (2) suppenemiseksi käytettävä hyvin pientä h:n arvoa. Tällöin Newtonin menetelmä on yleensä tehokkaampi.

Muutamia menetelmiä mainitaksemme Adams-Bashfort on eksplisiittinen ennustajamenetelmä kun taas Adams-Moulton on implisiittinen korjaajamenetelmä. Lisäksi esimerkiksi seuraavalla ennustaja-korjaaja-menetelmällä on globaalinen katkaisuvirhe $O(h^4)$:

$$y_{n+4} = y_{n+3} + \frac{h}{24}(55f_{n+3} - 59f_{n+2} + 37f_{n+1} - 9f_n)$$
$$y_{n+4} = y_{n+3} + \frac{h}{24}(9f_{n+4} + 19f_{n+3} - 5f_{n+2} + f_{n+1})$$

5.4 Reuna-arvotehtävät

Tarkastellaan toisen kertaluvun differentiaaliyhtälöön liittyvää *reuna-arvotehtävää*

$$\begin{cases} y'' = f(x, y, y') \\ y(a) = \alpha \\ y(b) = \beta \end{cases}$$

missä f on annettu funktio. Tehtävässä haetaan sellaista differentiaaliyhtälön välillä $[a, b]$ määriteltyä ratkaisua, jolla on annetut arvot välin reunapisteissä. Ehtoja $y(a) = \alpha$ ja $y(b) = \beta$ kutsutaan *reunaehdoiksi.*

Esimerkki. Yksiulotteisen lämpöyhtälön stationaarinen tapaus

$$\frac{d}{dx}\left(k(x)\frac{dy}{dx}\right) = 0$$

reunaehdoin $y(a) = \alpha$, $y(b) = \beta$ kuvaa lämpötilaa x-akselin välillä $[a, b]$ sijaitsevassa eristetyssä tangossa, jonka lämpötila $y(x)$ pisteessä x on ajan suhteen vakio. Jos lämmönjohtuvuudesta riippuva funktio $k(x)$ on vakio, reuna-arvotehtävän ratkaisu on 1. asteen polynomi. Jos $k(x)$ ei ole vakio, tehtävä joudutaan yleensä ratkaisemaan numeerisesti.

Tähtäysmenetelmä

Tarkastellaan reuna-arvotehtävän

$$\begin{cases} y'' = f(x, y, y') \\ y(a) = \alpha \\ y(b) = \beta \end{cases} \tag{1}$$

ohella alkuarvotehtävää

$$\begin{cases} y'' = f(x, y, y') \\ y(a) = \alpha \\ y'(a) = \gamma \end{cases} \tag{2}$$

ja oletetaan, että kummallakin tehtävällä on yksikäsitteinen ratkaisu. Tähtäysmenetelmässä pyritään löytämään γ siten, että alkuarvotehtävän (2) ratkaisu toteuttaa reunaehdon $y(b) = \beta$. Tällöin reuna-arvotehtävä (1) palautuu alkuarvotehtäväksi (2). Kun γ on löydetty, (2) voidaan esittää ensimmäisen kertaluvun systeeminä

$$\begin{cases} y' = v \\ v' = f(x, y, v) \\ y(a) = \alpha \\ v(a) = \gamma \end{cases} \tag{3}$$

Tämä voidaan ratkaista esimerkiksi Runge-Kutta-menetelmällä. γ löytyy parantamalla sopivaa alkuarvausta iteroimalla.

Esimerkki. Reuna-arvotehtävän

$$y'' = -y, \qquad y(0) = 0, \quad y\left(\frac{\pi}{2}\right) = 1$$

analyyttinen ratkaisu on $y(x) = \sin x$. Jos ratkaisua approksimoidaan reunaehdot toteuttavalla lineaarisella interpolointipolynomilla, saadaan γ:lle alkuarvaukseksi vastaavan suoran kulmakerroin

$$\gamma_0 = \frac{1-0}{\frac{\pi}{2}-0} = \frac{2}{\pi} = 0.637.$$

Alkuarvotehtävä (2) on nyt

$$\begin{cases} y'' = -y \\ y(0) = 0 \\ y'(0) = \gamma_0 = 0.637 \end{cases}$$

ja systeemi (3) on nyt

$$\begin{cases} y' = v \\ v' = -y \\ y(0) = 0 \\ v(0) = 0.637 \end{cases}$$

Runge-Kutta menetelmällä ($h = \frac{\pi}{100}$) saadulla ratkaisulla pätee $y(\frac{\pi}{2}) = 0.637$. Tähdättäessä γ_0:n suuntaan osuttiin siis maalin ($y(\frac{\pi}{2}) = 1$) alapuolelle. Muutetaan tähtäyssuuntaa ylemmäs ja valitaan esimerkiksi $y'(0) = \gamma_1 = 1.2$. Ratkaisemalla systeemi (3) kuten yllä saadaan tällöin $y(\frac{\pi}{2}) = 1.2$.

Yritys-erehdysmenetelmän asemasta voidaan oikeaa γ:n arvoa etsiä myös iteratiivisesti. Olkoon $y(x, \gamma)$ alkuarvotehtävän (2) ratkaisu γ:n funktiona ja merkitään

$$g(\gamma) = y(b, \gamma).$$

Tällöin $y(x, \gamma)$ on (1):n ratkaisu täsmälleen silloin kun $g(\gamma) = \beta$. γ:n määräämiseksi on ratkaistava yhtälö

$$g(\gamma) - \beta = 0. \tag{4}$$

Koska g:llä ei ole eksplisiittistä lauseketta, g:n arvot on laskettava numeerisesti. Myös g:n derivointi on hankalaa, joten yhtälöä (4) voidaan yrittää ratkaista *sekanttimenetelmällä*

$$\gamma_{n+1} = \gamma_n - \frac{g(\gamma_n) - \beta}{\dfrac{g(\gamma_n) - g(\gamma_{n-1})}{\gamma_n - \gamma_{n-1}}} \tag{5}$$

Esimerkki. Edellisen esimerkin alkuarvotehtävälle laskettiin

$$g(\gamma_0) = 0.637 \quad \text{ja} \quad g(\gamma_1) = 1.2,$$

missä $\gamma_0 = 0.637$ ja $\gamma_1 = 1.2$. Sekanttimenetelmällä saadaan ($n = 1$)

$$\gamma_2 = 1.2 - \frac{1.2 - 1}{\dfrac{1.2 - 0.637}{1.2 - 0.637}} = 1.2 - 0.2 = 1$$

Iterointia jatkettaessa huomataan, että $g(\gamma_2) = y(\frac{\pi}{2}, 1) = 1.0000$; nyt osuttiin siis suoraan maaliin ja yhtälön (4) ratkaisu on $\gamma_2 = 1$.

Voidaan näyttää, että funktion f riippuessa lineaarisesti y:stä ja y':sta kuten yllä olevassa esimerkissä, $g(\gamma)$ on astetta 1 oleva γ:n polynomi. Tästä johtuen sekanttimenetelmä antoi heti tarkan ratkaisun. Differentiaaliyhtälön ollessa epälineaarinen joudutaan yhtälöä (4) ratkaistaessa yleensä suorittamaan useampia iteraatioita.

Esimerkki. Ratkaistaan reuna-arvotehtävä

$$y'' = 1 + yy', \qquad y(0) = 1, \quad y(0.6) = 2$$

tähtäysmenetelmällä käyttämällä alkuarvotehtävissä Runge-Kutta-menetelmää askelpituudella $h = 0.01$. Lineaarisella interpoloinnilla saadaan alkuarvaus

$$\gamma_0 = \frac{2 - 1}{0.6 - 0} = 1.67$$

jolloin $g(\gamma_0) = 2.8788$. Tähtäämällä alemmas ($\gamma_1 = 0.8$) tulee $g(\gamma_1) = 1.3544$. Sekanttimenetelmällä saadaan $\gamma_2 = 0.8429$; tällöin $g(\gamma_2) = 1.9965$. Edelleen $\gamma_3 = 0.8465$ ja $g(\gamma_3) = 2.0000$ on jo varsin lähellä oikeaa reuna-arvoa.