

Matti Salán

ALKUARVOTEHTÄVIEN RATKAISEMINEN NUMEERISESTI

Kandidaatintyö
Tekniikan ja luonnontieteiden tiedekunta
Tarkastaja: Simo Ali-Löyty
Huhtikuu 2022

TIIVISTELMÄ

Matti Salán: Alkuarvotehtävien ratkaiseminen numeerisesti
Kandidaatintyö
Tampereen yliopisto
Tekniikka ja luonnontieteet, TkK
Huhtikuu 2022

Monia käytännön ilmiötä voidaan mallintaa matemaattisesti differentiaaliyhtälöiden avulla. Tällaisia ilmiöitä ovat esimerkiksi tarttuvan taudin leviäminen ja heilurin liike. Alkuarvotehtävät ovat sellaisia differentiaaliyhtälöitä, joiden ratkaisuille annetaan jokin alkuehto. Useissa tapauksissa differentiaaliyhtälöihin perustuvat mallit ovat kuitenkin sellaisia, että niiden analyttinen ratkaiseminen on hankalaa tai jopa mahdotonta. Tällöin voidaan turvautua alkuarvotehtävien numeeriseen ratkaisuun. Tässä työssä tutkitaan erilaisia numeerisia menetelmiä, joilla voidaan ratkaista alkuarvotehtäviä numeerisesti. Työn tavoite on esitellä yksiaskelmenetelmiä, ja vertailla eri yksiaskelmenetelmiä keskenään.

Työ jakautuu teoriaosaan ja käytännön toteutukseen. Teoriaosassa esitellään ensin yksiaskelmenetelmiä yleisellä tasolla. Tämän jälkeen tutkitaan Eulerin menetelmää, Taylorin menetelmiä ja Runge–Kutta-menetelmiä tarkemmin. Teoriaosan jälkeen ratkaistaan alkuarvotehtävä numeerisesti MATLAB-ohjelmiston avulla. Lisäksi vertaillaan eri menetelmiä keskenään.

Työssä näytetään, että Eulerin menetelmän kertaluku on yksi. Tämän lisäksi saadaan yläraja Eulerin menetelmän kokonaisvirheen itseisarvolle, minkä perusteella Eulerin menetelmä on supeneva. Ratkaistaan myös 2-vaiheisen Runge–Kutta-menetelmän kertoimet, ja nähdään, että 2-vaiheisia 2. kertaluvun Runge–Kutta-menetelmiä on ääretön määrä. Työssä esitellään yksi tällainen menetelmä sekä 4. kertaluvun Klassinen Runge–Kutta-menetelmä, joka on tunnetuin eksplisiittinen Runge–Kutta-menetelmä.

Käytännön toteutuksessa muodostetaan alkuarvotehtävä Newtonin 2. lain avulla ja ratkaistaan se kolmella eri Runge–Kutta-menetelmällä. Lisäksi vertaillaan numeerisia ratkaisuja alkuarvotehtävän tarkkaan ratkaisuun. Tämä vertailu tehtiin laskemalla kokonaisvirheiden itseisarvoja. Vertailulla havaitaan, että korkeamman kertaluvun menetelmät lähestyvät tarkkaa ratkaisua nopeammin kuin Eulerin menetelmä, kun askelpituutta pienennetään.

Avainsanat: alkuarvotehtävä, numeeriset menetelmät, Runge–Kutta-menetelmät, Eulerin menetelmä

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

SISÄLLYSLUETTELO

| | | |
|-----|---|----|
| 1. | Johdanto | 1 |
| 2. | Alkuarvotehtävät | 2 |
| 3. | Numeeriset menetelmät | 4 |
| 3.1 | Yksiaskelmenetelmät ja määritelmiä | 4 |
| 3.2 | Eulerin menetelmä | 7 |
| 3.3 | Taylorin menetelmät | 8 |
| 3.4 | Eulerin menetelmän analyysi | 10 |
| 4. | Runge–Kutta-menetelmät | 15 |
| 4.1 | Eksplisiittiset ja implisiittiset Runge–Kutta-menetelmät | 15 |
| 4.2 | 2-vaiheinen Runge–Kutta-menetelmä | 17 |
| 5. | Runge–Kutta-menetelmät käytännössä | 21 |
| 5.1 | Alkuarvotehtävän numeerinen ratkaisu: putoava pallo | 21 |
| 5.2 | Menetelmien vertailu | 23 |
| 5.3 | Vektorimuotoisen alkuarvotehtävän ratkaiseminen | 25 |
| 6. | Yhteenveto | 27 |
| | Lähteet. | 28 |
| | Liite A: MATLAB-funktiot numeerisille menetelmille | 29 |
| A.1 | Eulerin menetelmä | 29 |
| A.2 | Ralstonin menetelmä | 29 |
| A.3 | Klassinen Runge–Kutta-menetelmä | 30 |
| | Liite B: Alkuarvotehtävien ratkaisemisessa käytetyt MATLAB-koodit | 31 |
| B.1 | Numeerinen ratkaiseminen Eulerin menetelmällä | 31 |
| B.2 | Alkuarvotehtävän ratkaiseminen: putoava pallo | 32 |
| B.3 | Kokonaisvirheen itseisarvon maksimi eri askelpituuksilla | 33 |
| B.4 | Vektorimuotoisen alkuarvotehtävän ratkaisu. | 34 |

LYHENTEET JA MERKINNÄT

| | |
|--------------------------|---|
| A | matriisi |
| L | funktion $f(x, y)$ Lipschitz-vakio |
| N | askelten määrä |
| $[a, b]$ | suljettu reaalilukuväli |
| Φ | yksiaskelmenetelmän lisäysfunktio |
| \mathbf{b}^\top | vektorin \mathbf{b} transpoosi |
| \mathbf{b}, \mathbf{c} | vektoreita |
| \mathbf{f}, \mathbf{y} | vektoriarvoisia funktiota |
| $\ \cdot\ $ | vektorin euklidinen normi |
| $ \cdot $ | itseisarvo |
| \mathbb{R}^n | n -ulotteinen reaaliavaruus |
| \mathbb{R} | reaalilukujen joukko |
| \subset | osajoukko |
| τ_n | paikallinen diskretointivirhe |
| \times | tulojoukko |
| $\{a_n\}$ | lukujono |
| e_n | kokonaisvirhe |
| f' | funktion f derivaatta muuttujan x suhteen |
| $f^{(k)}$ | funktion f k :n kertaluvun derivaatta muuttujan x suhteen |
| f_x, f_y | funktion f osittaisderivaatta muuttujan x tai muuttujan y suhteen |
| f_{yx} | funktion f toisen kertaluvun osittaisderivaatta |
| f | yhden tai usean muuttujan reaaliarvoinen funktio |
| h | askelpituus |
| p | numeerisen menetelmän kertaluku |
| y_0 | alkuarvotettävän alkuarvo ja numeerinen ratkaisu pisteessä x_0 |
| $y(x_n)$ | tarkka ratkaisu pisteessä x_n |
| y_n | numeerinen ratkaisu pisteessä x_n |

1. JOHDANTO

Tavallisella differentiaaliyhtälöllä tarkoitetaan yhtälöä, joka sisältää tuntemattoman funktion ja sen derivaattoja [4, s. 1]. Kun ratkaistaan tällaista yhtälöä, niin tarkoituksena on löytää funktiot, jotka toteuttavat differentiaaliyhtälön. Alkuarvotehtävät ovat tavallisia differentiaaliyhtälöitä, mutta differentiaaliyhtälön ratkaisulle annetaan niin sanottuja alkuehtoja, jotka ratkaisun tulee toteuttaa. Alkuarvotehtävät tulevat usein vastaan, kun mallinnetaan käytännön ilmiöitä matemaattisesti. Usein nämä mallit ovat kuitenkin sellaisia, että alkuarvotehtävän analyttisen ratkaisun löytäminen on hankalaa tai jopa mahdotonta [2, s. 3]. Tällöin voidaan turvautua numeerisiin menetelmiin.

Tämän työn tarkoituksena on esitellä alkuarvotehtävien numeerista ratkaisemista yksiaskelmenetelmillä ja vertailla eri menetelmiä keskenään. Teoreettisen tarkastelun lisäksi tarkoituksena on ratkaista alkuarvotehtävä numeerisesti MATLAB-ohjelmistoa hyödyntäen.

Yksinkertaisin numeerinen menetelmä alkuarvotehtävien ratkaisemiseen on niin sanottu Eulerin menetelmä, joka on nimetty kuuluisan Leonhard Eulerin (1707–1783) mukaan [3, s. 341]. Eulerin menetelmä ei kuitenkaan ole laskennallisesti kauhean tehokas, sillä jos halutaan hyvä approksimaatio, niin se vaatii usein paljon laskentaa verrattuna muihin menetelmiin. Eulerin menetelmä on niin sanottu eksplisiittinen yksiaskelmenetelmä. Lisäksi Eulerin menetelmä on yksinkertaisin Runge–Kutta-menetelmä, jotka ovat tunnetuimpia yksiaskelmenetelmiä [3, s. 418].

Tässä työssä tarkastellaan ensin alkuarvotehtävien numeerista ratkaisemista yleisesti, minkä jälkeen tarkastellaan muutamaa numeerista menetelmää tarkemmin. Luvussa 2 kerrotaan, mitä alkuarvotehtävät ovat, ja milloin alkuarvotehtävällä on olemassa yksikäsitteinen ratkaisu. Tämän jälkeen Luvussa 3 lähdetään tarkastelemaan alkuarvotehtävien ratkaisemista numeerisesti. Ensin tarkastellaan yksiaskelmenetelmiä yleisellä tasolla, minkä jälkeen perehdytään Eulerin menetelmään ja Taylorin menetelmiin, jotka ovat eksplisiittisiä yksiaskelmenetelmiä. Luvussa 4 esitellään Runge–Kutta-menetelmät ja johdetaan eksplisiittisen 2-vaiheisen Runge–Kutta-menetelmän kertoimet. Lopuksi Luvussa 5 ratkaistaan alkuarvotehtävä kolmella eri Runge–Kutta-menetelmällä ja vertaillaan näitä menetelmiä keskenään.

2. ALKUARVOTEHTÄVÄT

Tässä työssä tarkastellaan erityisesti 1. kertaluvun alkuarvotehtäviä ja niiden ratkaisemista numeerisesti. Tällainen alkuarvotehtävä on muotoa

$$y'(x) = f(x, y(x)), \quad y(x_0) = y_0, \quad (2.1)$$

missä $x \in [a, b] \subset \mathbb{R}$ ja x_0 on jokin kyseisen välin piste. Vaaditaan lisäksi, että funktio f on jatkuva jokaisella (x, y) joukossa $D \subset \mathbb{R}^2$. Jos jokaisella $x \in [a, b]$

$$\begin{aligned} (x, y(x)) &\in D, \\ y(x_0) &= y_0, \\ y'(x) &\text{ on olemassa ja } y'(x) = f(x, y(x)), \end{aligned} \quad (2.2)$$

niin funktio $y(x)$ on alkuarvotehtävän (2.1) ratkaisu välillä $[a, b]$. [3, s. 333] Usein tuntemattoman funktion y argumentti x jätetään kirjoittamatta, eli differentiaaliyhtälö (2.1) kirjoitetaan muodossa $y' = f(x, y)$.

Jos alkuarvotehtävä, jonka kertaluku on n , voidaan kirjoittaa muotoon

$$\begin{aligned} y^{(n)} &= f(x, y, y', y'', \dots, y^{(n-1)}), \\ y(x_0) &= z_1, y'(x_0) = z_2, \dots, y^{(n-1)}(x_0) = z_n \end{aligned} \quad (2.3)$$

niin se voidaan muuntaa 1. kertaluvun differentiaaliyhtälöiden ryhmäksi. Tämä onnistuu ottamalla käyttöön funktiot

$$y_1 = y, \quad y_2 = y', \quad \dots, \quad y_n = y^{(n-1)}, \quad (2.4)$$

jolloin alkuarvotehtävä (2.3) saadaan muotoon

$$\begin{aligned} y_1' &= y_2, & y_1(x_0) &= z_1, \\ y_2' &= y_3, & y_2(x_0) &= z_2, \\ &\vdots & &\vdots \\ y_n' &= f(x, y_1, y_2, \dots, y_n), & y_n(x_0) &= z_n. \end{aligned} \quad (2.5)$$

Merkitään

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ f(x, y_1, y_2, \dots, y_n) \end{bmatrix}, \quad \mathbf{y}_0 = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}, \quad (2.6)$$

jolloin (2.5) saadaan muotoon

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{y}_0. \quad (2.7)$$

Saatu muoto on 1. kertaluvun alkuarvotehtävää (2.1) vastaavassa muodossa, eli tutkielmassa esitettävät menetelmät yleistyvät suoraviivaisesti korkeamman kertaluvun alkuarvotehtävien ratkaisemiseen. [3, s. 340] Täten vain 1. kertaluvun alkuarvotehtävien tarkastelu ei ole rajoitus.

Käydään seuraavaksi läpi, milloin alkuarvotehtävällä on olemassa yksikäsitteinen ratkaisu. Tätä varten tarvitaan Lipschitz-jatkuvuuden käsite.

Määritelmä 2.1. Funktio $f : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ on *Lipschitz-jatkuva* muuttujan \mathbf{y} suhteen, jos on olemassa sellainen vakio $L > 0$, että

$$\|\mathbf{f}(x, \mathbf{u}) - \mathbf{f}(x, \mathbf{v})\| \leq L\|\mathbf{u} - \mathbf{v}\| \quad (2.8)$$

jokaisella $x \in [a, b]$ ja $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Ehtoa (2.8) sanotaan *Lipschitz-ehdoksi* ja vakiota L sanotaan *Lipschitz-vakioksi*. [6, s. 23]

Vastaavasti määritellään funktion $f : [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$ Lipschitz-jatkuvuus, mutta vektorinormit korvataan itseisarvoilla. Nyt voidaan esittää lause, joka kertoo, milloin alkuarvotehtävällä on olemassa yksikäsitteinen ratkaisu.

Lause 2.2. *Tarkastellaan alkuarvotehtävää*

$$\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}(x)), \quad \mathbf{y}(a) = \mathbf{y}_0. \quad (2.9)$$

Jos funktio $f : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ on jatkuva ensimmäisen muuttujansa suhteen ja toteuttaa Lipschitz-ehdon toisen muuttujansa suhteen, niin alkuarvotehtävälle (2.9) on olemassa yksikäsitteinen ratkaisu.

Edellisen lauseen todistamiseen tarvitaan metrisen avaruuden käsite sekä niin sanottu kontraktio-kuvausperiaate, joten sivuutetaan lauseen todistus. Todistus löytyy kirjasta [6, s. 23–24].

3. NUMEERISET MENETELMÄT

Tarkastellaan tästä eteenpäin alkuarvotehtävän (2.1) numeerista ratkaisemista välillä $[a, b]$, kun valitaan $x_0 = a$. Numeerisessa ratkaisemisessa lähdetään liikkeelle siitä, että tarkasteltava väli jaetaan $N + 1$ pisteeseen seuraavasti:

$$a = x_0 < x_1 < \dots < x_{N-1} < x_N = b. \quad (3.1)$$

Kahden peräkkäisen pisteen välistä erotusta $h_n = x_{n+1} - x_n$ kutsutaan askelpituudeksi. [5, s. 537] Pidetään jatkossa askelpituus vakiona

$$h = \frac{b - a}{N}. \quad (3.2)$$

Alkuarvotehtävän tarkka ratkaisu on siis jatkuva funktio tarkasteltavalla välillä $[a, b]$, mutta numeerisessa ratkaisemisessa arvioidaan alkuarvotehtävän tarkkaa ratkaisua pisteissä (3.1) [5, s. 537]. Jatkossa merkintä y_n tarkoittaa alkuarvotehtävän numeerista ratkaisua pisteessä x_n , $n = 0, 1, \dots, N$. Numeerinen ratkaisu y_0 on suoraan alkuarvotehtävän alkuarvo, minkä avulla saadaan numeeriset ratkaisut muissa pisteissä. Numeeriset menetelmät voidaan jakaa eri kategorioihin, ja tässä tutkielmassa tarkastellaan erityisesti eksplisiittisiä yksiaskelmenetelmiä.

Jatkossa tullaan käyttämään merkintää $z = O(h^p)$, missä p on jokin positiivinen kokonaisluku. Tällä merkinnällä tarkoitetaan sitä, että z vähenee vähintään yhtä nopeasti kuin h^p , kun h on tarpeeksi pieni. [8, s. 22] Määritellään edellinen merkintätapa täsmällisemmin.

Määritelmä 3.1. Jos on olemassa sellaiset positiiviset luvut c ja h_0 , että

$$|z| \leq ch^p \quad (3.3)$$

jokaisella $0 < h < h_0$, niin merkitään $z = O(h^p)$. [8, s. 23]

3.1 Yksiaskelmenetelmät ja määritelmiä

Nimitys yksiaskelmenetelmille tulee siitä, että numeerinen ratkaisu y_{n+1} saadaan edellisen numeerisen ratkaisun y_n avulla. Moniaskelmenetelmissä numeerinen ratkaisu y_{n+1} saadaan taas useamman aikaisemman numeerisen ratkaisun avulla, esimerkiksi kaksiaskelmenetelmillä numeerinen ratkaisu y_{n+1} saadaan ratkaisujen y_n ja y_{n-1} avulla [5, s. 538–539].

Määritelmä 3.2. Yleinen muoto *yksiaskelmenetelmälle* on

$$y_{n+1} = y_n + h\Phi(x_n, y_n, y_{n+1}, h), \quad n = 0, 1, \dots, N-1, \quad (3.4)$$

missä $h = x_{n+1} - x_n$ on *askelpituus*, ja *lisäysfunktio* Φ on jatkuva funktio muuttujiensa suhteen. Jos lisäysfunktio Φ on riippumaton muuttujasta y_{n+1} , eli

$$y_{n+1} = y_n + h\Phi(x_n, y_n, h), \quad n = 0, 1, \dots, N-1, \quad (3.5)$$

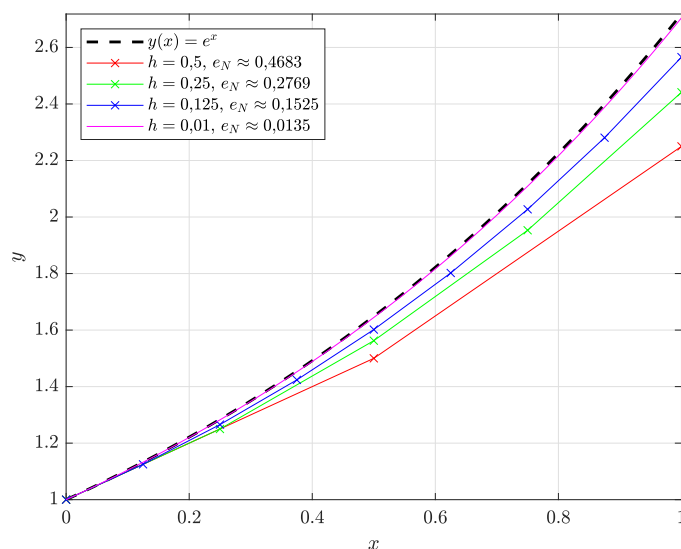
niin yksiaskelmenetelmää sanotaan *eksplisiittiseksi*. Jos taas lisäysfunktio Φ riippuu muuttujasta y_{n+1} , niin yksiaskelmenetelmää kutsutaan *implisiittiseksi*. [5, s. 538]

Alkuarvotehtävän (2.1) numeerisessa ratkaisussa yksiaskelmenetelmän avulla on siis kyse siitä, että lähdetään liikkeelle pisteestä $x_0 = a$, jossa ratkaisu y_0 saadaan suoraan alkuehdosta. Tämän avulla saadaan numeerinen ratkaisu pisteessä x_1 , jonka avulla saadaan ratkaisu pisteessä x_2 . Näin edetään aina pisteeseen $x_N = b$ asti. Numeerisen menetelmän lisäysfunktio Φ kuvaa alkuarvotehtävän ratkaisun suuntaa, ja jokaisella askeleella siirrytään tähän suuntaan askelpituuden verran. Havainnollistetaan yksiaskelmenetelmien toimintaa seuraavan esimerkin avulla.

Esimerkki 3.3. Ratkaistaan alkuarvotehtävä

$$y' = y, \quad y(0) = 1 \quad (3.6)$$

numeerisesti välillä $0 \leq x \leq 1$. Käytetään niin sanottua Eulerin menetelmää, joka on eksplisiittinen yksiaskelmenetelmä. Funktio $y(x) = e^x$ on tämän alkuarvotehtävän tarkka ratkaisu, sillä $y'(x) = e^x$ ja $y(0) = e^0 = 1$. Kuvassa 3.1 on esitetty tarkka ratkaisu sekä numeeriset ratkaisut eri askelpituuksilla. Kuva 3.1 saatiin Liitteen B.1 MATLAB-koodilla. Kuviin piirretään usein



Kuva 3.1. Alkuarvotehtävän (3.6) ratkaisu ja numeeriset ratkaisut eri askelpituuksilla.

murtoviiva, joka kulkee numeerisen menetelmän avulla laskettujen pisteiden kautta. Kuvasta 3.1

nähdään, että numeerinen ratkaisu lähestyy tarkkaa ratkaisua, kun askelpituutta pienennetään. Tämä nähdään jo selvästi kuvasta, mutta myös kokonaisvirheistä e_N , jotka on laskettu kohdassa $x = 1$, ks. Määritelmä 3.4. Askelpituudella $h = 0,5$ kokonaisvirhe kohdassa $x = 1$ on noin 0,4683, mutta askelpituudella $h = 0,01$ vastaava virhe on enää noin 0,0135.

Numeerisesta menetelmästä ei ole kauheasti hyötyä, jos se ei anna tarpeeksi tarkkaa ratkaisua. Seuraavassa määritelmässä esitetään kaksi keskeistä käsitettä, joiden avulla voidaan arvioida numeerisen menetelmän tarkkuutta.

Määritelmä 3.4. Määritellään

$$\tau_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n), h). \quad (3.7)$$

Tätä kutsutaan *yksiaskelmenetelmän paikalliseksi diskreetointivirheeksi*. Se kuvaa virhettä, joka tehdään yhdellä menetelmän askeleella, kun oletetaan, että edelliset ratkaisut ovat tarkkoja. Paikallista diskreetointivirhettä olennaisempi virhe on *kokonaisvirhe*

$$e_n = y(x_n) - y_n. \quad (3.8)$$

Kokonaisvirhe kertoo siis, kuinka paljon tarkka ratkaisu pisteessä x_n eroaa numeerisesta ratkaisusta samassa pisteessä. [5, s. 540]

Kokonaisvirheen voisi ajatella kasvavan jokaisella askeleella, mutta tämä ei kuitenkaan aina ole totta [5, s. 540]. Jos esimerkiksi ensimmäisellä askelleella kokonaisvirhe kasvaa, niin seuraavalla askeleella voidaan mahdollisesti päästä lähemmäs tarkkaa ratkaisua. Tällöin kokonaisvirhe pienentyisi. Esimerkki tällaisesta käyttäytymisestä nähdään Luvussa 5, ks. Kuva 5.2a.

Yksittäisten kokonaisvirheiden tarkastelu ei siis kerro kauheasti menetelmän tarkkuudesta. Menetelmän tarkkuutta voidaan kuitenkin tarkastella kokonaisvirheen itseisarvon maksimiarvon avulla. Mitä pienempi kokonaisvirheen itseisarvon maksimi, sitä tarkempi numeerinen ratkaisu on. Jos siis kokonaisvirheen itseisarvon maksimi lähestyy nollaa, niin numeerisen menetelmän antama ratkaisu lähestyy tarkkaa ratkaisua.

Määritelmä 3.5. Jos

$$\lim_{h \rightarrow 0} \max_{1 \leq n \leq N} |e_n| = 0, \quad (3.9)$$

niin sanotaan, että numeerinen menetelmä on *suppeneva* [5, s. 541].

Ollaan siis kiinnostuneita siitä, miten kokonaisvirhe käyttäytyy, kun askelpituutta pienennetään.

Kaksi muuta tärkeää käsitettä ovat konsistenssi ja numeerisen menetelmän kertaluku, jotka määritellään seuraavaksi.

Määritelmä 3.6. Numeerinen menetelmä (3.5) on *konsistenssi*, jos

$$\lim_{h \rightarrow 0} \tau_n = 0 \quad (3.10)$$

Lisäksi jos

$$\tau_n = O(h^p), \quad (3.11)$$

missä p on jokin positiivinen kokonaisluku, niin *numeerisen menetelmän kertaluku* on p . [5, s. 540]

Otetaan paikallisesta diskreetointivirheestä (3.7) raja-arvo, kun $h \rightarrow 0$. Määritelmässä 3.2 oletettiin, että lisäysfunktio Φ on jatkuva kaikkien muuttujiensa suhteen. Lisäksi y' on jatkuva, joten

$$\lim_{h \rightarrow 0} \tau_n = y'(x_n) - \Phi(x_n, y(x_n), 0) = f(x_n, y(x_n)) - \Phi(x_n, y(x_n), 0). \quad (3.12)$$

Tästä seuraa, että yksiaskelmenetelmä on konsistenssi, jos

$$f(x_n, y(x_n)) = \Phi(x_n, y(x_n), 0). \quad (3.13)$$

Ehtoa (3.13) pidetään joskus konsistenssin määritelmänä. [11, s. 322]

Edellä määriteltiin kaksi tärkeää numeeristen menetelmien käsitettä, jotka olivat konsistenssi ja suppeneminen. Kolmas tärkeä käsite on numeerisen menetelmän stabiilisuus, mutta sitä ei käsitellä tässä tutkielmassa. Näistä kolmesta käsitteestä tärkein on kuitenkin numeerisen menetelmän suppeneminen, sillä numeerisesta menetelmästä ei ole käytännössä mitään hyötyä, jos menetelmän antamat ratkaisut eivät lähesty alkuarvottehtävän tarkkaa ratkaisua, kun askelpituutta pienennetään [5, s. 541].

3.2 Eulerin menetelmä

Yksinkertaisin numeerinen menetelmä alkuarvottehtävän numeeriseen ratkaisemiseen on niin sanottu Eulerin menetelmä, joka on eksplisiittinen yksiaskelmenetelmä. Eulerin menetelmä on

$$y_{n+1} = y_n + hf(x_n, y_n), \quad n = 0, 1, \dots, N-1, \quad (3.14)$$

missä y_0 on tarkasteltavan alkuarvottehtävän alkuarvo ja funktio f on alkuarvottehtävässä esiintyvä funktio, ks. (2.1). Nyt siis valitsemalla järjestyksessä $n = 0, 1, \dots, N-1$, saadaan laskettua approksimaatiot y_1, y_2, \dots, y_N . [11, s. 317]

Eulerin menetelmä voidaan johtaa usealla eri tavalla. Käydään läpi kaksi eri tapaa. Kirjoitetaan ensin differentiaaliyhtälö (2.1) pisteessä x_n , eli

$$y'(x_n) = f(x_n, y(x_n)). \quad (3.15)$$

Funktion y derivaattaa pisteessä x_n voidaan approksimoida oikeanpuoleisen erotusosamäärän avulla, joten

$$\frac{y(x_{n+1}) - y(x_n)}{h} \approx f(x_n, y(x_n)). \quad (3.16)$$

Nyt jos $y(x_{n+1})$ ja $y(x_n)$ korvataan niitä vastaavilla numeerisilla approksimaatiolla y_{n+1} ja y_n , niin saadaan Eulerin menetelmä (3.14). [4, s. 419]

Toinen tapa johtaa Eulerin menetelmä on differentiaaliyhtälön (2.1) integrointi puolittain pisteestä

x_n pisteeseen x_{n+1} , eli

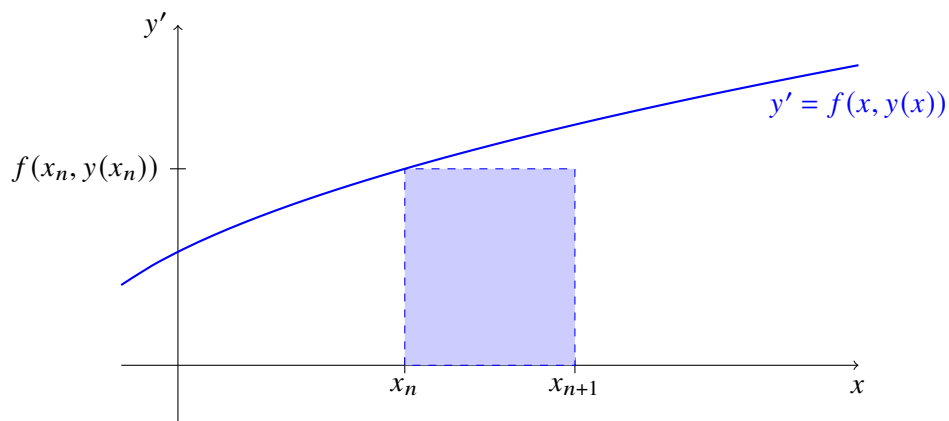
$$\int_{x_n}^{x_{n+1}} y'(x) dx = \int_{x_n}^{x_{n+1}} f(x, y(x)) dx. \quad (3.17)$$

Laskemalla yhtälön (3.17) vasemman puolen integraali, saadaan

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x)) dx. \quad (3.18)$$

Approksimoidaan yhtälön (3.18) integraalia approksimoidaan Kuvan 3.2 suorakulmion pinta-alalla, eli

$$\int_{x_n}^{x_{n+1}} f(x, y(x)) dx \approx (x_{n+1} - x_n) f(x_n, y(x_n)) = hf(x_n, y(x_n)). \quad (3.19)$$



Kuva 3.2. Integraalin approksimointi suorakulmion pinta-alan avulla, perustuu lähteeseen [4, s. 420]

Tällöin

$$y(x_{n+1}) \approx y(x_n) + hf(x_n, y(x_n)), \quad (3.20)$$

ja edelleen korvaamalla $y(x_{n+1})$ ja $y(x_n)$ vastaavilla numeerisilla approksimaatiolla y_{n+1} ja y_n , saadaan Eulerin menetelmä (3.14). [4, s. 419] Eulerin menetelmä voidaan johtaa myös geometrisesti ja Taylorin sarjan avulla [3, s. 342], joista jälkimmäinen tehdään seuraavassa alaluvussa.

3.3 Taylorin menetelmät

Toinen tapa ratkaista alkuarvot tehtävä (2.1) numeerisesti on niin sanottu Taylorin menetelmä, joka perustuu seuraavaan määritelmään.

Määritelmä 3.7. Jos funktiolla $f(x)$ on kaikkien kertalukujen derivaatat pisteessä $x = c$, niin sarjaa

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!} (x-c)^k = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!} (x-c)^2 + \dots \quad (3.21)$$

kutsutaan funktion f Taylorin sarjaksi pisteen c suhteen. [1, s. 513]

Edellinen määritelmä ei vaadi, että sarja (3.21) suppenee muissa pisteissä kuin $x = c$, jossa sarja on vain $f(c) + 0 + 0 + \dots$. Sarja (3.21) on määritelty, kun funktion f kaikki derivaatat ovat olemassa pisteessä $x = c$. Tämä tarkoittaa sitä, että derivaatat tulee olla olemassa avoimella välillä, johon luku c kuuluu. Sarja ei kuitenkaan välttämättä suppene missään muussa pisteessä kuin $x = c$. [1, s. 513]

Funktiota $f(x)$ voidaan approksimoida Taylorin sarjan (3.21) osasumman avulla. Taylorin sarjan n :ttä osasummaa

$$P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(c)}{k!} (x-c)^k \quad (3.22)$$

kutsutaan n . asteen Taylorin polynomiksi pisteen c suhteen [1, s. 257]. Seuraava lause antaa virheen, joka tehdään approksimoidessa funktiota f Taylorin polynomilla, eli $f(x) \approx P_n(x)$.

Lause 3.8. *Olkoon funktiolla $f(x)$ olemassa kertaluvun $n + 1$ derivaatta sellaisella välillä, johon kuuluu luvut c ja x . Tällöin approksimaatiossa $f(x) \approx P_n(x)$ tehty virhe*

$$E_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-c)^{(n+1)}, \quad (3.23)$$

missä ξ on jokin luku lukujen c ja x väliltä. Tällöin saadaan kaava

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(c)}{k!} (x-c)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-c)^{(n+1)}, \quad (3.24)$$

missä ξ on jokin luku lukujen c ja x väliltä. Kaavaa (3.24) kutsutaan Taylorin kaavaksi Lagrangen jäännöstermillä.

Sivuutetaan edellisen lauseen todistus, joka löytyy lähteestä [1, s. 259–260].

Nyt voidaan johtaa Eulerin menetelmä Taylorin kaavan avulla. Oletetaan, että alkuarvotehtävän ratkaisu on kaksi kertaa jatkuvasti derivoituva. Tällöin Taylorin kaavan (3.24) mukaan alkuarvotehtävän ratkaisu pisteessä x_{n+1} pisteen $c = x_n$ suhteen

$$\begin{aligned} y(x_{n+1}) &= y(x_n) + y'(x_n)(x_{n+1} - x_n) + \frac{y''(\xi_n)}{2}(x_{n+1} - x_n)^2 \\ &= y(x_n) + hf(x_n, y(x_n)) + \frac{h^2}{2}y''(\xi_n) \end{aligned} \quad (3.25)$$

missä $x_n < \xi_n < x_{n+1}$. Kun edellisestä yhtälöstä jätetään Lagrangen jäännöstermi pois, ja korvataan $y(x_{n+1})$ ja $y(x_n)$ vastaavilla numeerisilla approksimaatiolla, saadaan Eulerin menetelmä (3.14). [3, s. 342]

Eulerin menetelmä on yksinkertainen johtaa sekä käyttää, mutta se ei kuitenkaan ole kauhean tarkka. Seuraavassa alaluvussa nähdään, että Eulerin menetelmän kertaluku on 1. Eulerin menetelmä saatiin johdettua Taylorin kaavalla, mutta Taylorin kaavalla voidaan johtaa myös korkeamman kertaluvun menetelmiä. [5, s. 560]

Oletetaan, että alkuarvotehtävän (2.1) ratkaisu on $k + 1$ kertaa jatkuvasti derivoituva välillä $[a, b]$.

Käyttämällä Taylorin kaavaa pisteen x_n suhteen, saadaan

$$y(x_{n+1}) = y(x_n) + y'(x_n)h + \frac{y''(x_n)}{2}h^2 + \dots + \frac{y^{(k)}(x_n)}{k!}h^k + \frac{y^{(k+1)}(\xi_n)}{(k+1)!}h^{k+1}, \quad (3.26)$$

missä $x_n < \xi_n < x_{n+1}$. Tarkasteltavan differentiaaliyhtälön mukaan $y'(x_n) = f(x_n, y(x_n))$, ja korkeamman asteen derivaatat pisteessä x_n saadaan funktion $f(x, y(x))$ kokonaisderivaattojen avulla, jos ne ovat määriteltyjä pisteessä $(x_n, y(x_n))$. Nyt pudottamalla jäännöstermi yhtälöstä (3.26) ja korvaamalla $y(x_{n+1})$ ja $y(x_n)$ vastaavilla numeerisilla approksimaatiolla, saadaan

$$y_{n+1} = y_n + hf(x_n, y_n) + \frac{h^2}{2}f'(x_n, y_n) + \dots + \frac{h^k}{k!}f^{(k-1)}(x_n, y_n). \quad (3.27)$$

Tällaisia menetelmiä kutsutaan Taylorin menetelmiksi. [5, s. 560]

Taylorin menetelmät ovat eksplisiittisiä yksiaskelmenetelmiä, joiden lisäysfunktio

$$\Phi(x_n, y_n, h) = f(x_n, y_n) + \frac{h}{2}f'(x_n, y_n) + \dots + \frac{h^{k-1}}{k!}f^{(k-1)}(x_n, y_n). \quad (3.28)$$

Paikallinen diskreetointivirhe saadaan lisäysfunktion (3.28) ja yhtälön (3.26) avulla. Paikallinen diskreetointivirhe on siis

$$\tau_n = \frac{y^{(k+1)}(\xi_n)}{(k+1)!}h^k, \quad (3.29)$$

missä $x_n < \xi_n < x_{n+1}$. Edellisen yhtälön mukaan

$$|\tau_n| \leq \frac{h^k}{(k+1)!} \max_{x \in [a, b]} |y^{(k+1)}(x)|, \quad (3.30)$$

eli $\tau_n = O(h^k)$. Taylorin menetelmän (3.27) kertaluku on siis k [5, s. 561].

3.4 Eulerin menetelmän analyysi

Analysoidaan Eulerin menetelmää (3.14) hieman tarkemmin. Tarkastellaan paikallista diskreetointivirhettä sekä kokonaisvirhettä. Näiden avulla voidaan tutkia menetelmän suppenemista ja konsistenssia. Aloitetaan Eulerin menetelmän tarkastelu paikallisesta diskreetointivirheestä.

Edellisessä aluvussa johdettiin Eulerin menetelmä Taylorin sarjan avulla. Menetelmää johtaessa saatiin yhtälö (3.25), josta saadaan Eulerin menetelmän paikalliseksi diskreetointivirheeksi

$$\tau_n = \frac{y(x_{n+1}) - y(x_n)}{h} - f(x_n, y(x_n)) = \frac{h}{2}y''(\xi_n), \quad (3.31)$$

missä $x_n < \xi_n < x_{n+1}$. Jos differentiaaliyhtälön ratkaisu $y(x)$ on kaksi kertaa jatkuvasti derivoituva, niin paikallisen diskreetointivirheen itseisarvolle saadaan yläraja

$$|\tau_n| \leq \frac{h}{2} \max_{x \in [a, b]} |y''(x)|. \quad (3.32)$$

Tästä seuraa, että $\tau_n = O(h)$, eli Eulerin menetelmän kertaluku on 1. Lisäksi nähdään, että Eulerin

menetelmä on konsistenssi, sillä $\tau_n \rightarrow 0$, kun $h \rightarrow 0$. [5, s. 551]

Tarkastellaan seuraavaksi Eulerin menetelmän kokonaisvirhettä ja etsitään sen itseisarvolle yläraja. Tätä varten tarvitaan kaksi apulausetta.

Apulause 3.9. *Olkoon m jokin positiivinen reaaliluku ja $x \geq -1$. Tällöin*

$$0 \leq (1+x)^m \leq e^{mx}. \quad (3.33)$$

Todistus. (Vrt. [5, s. 552]) Taylorin kaavalla (3.24) pisteen $c = 0$ suhteen, saadaan

$$e^x = e^0 + e^0 x + \frac{1}{2} x^2 e^\xi = 1 + x + \frac{1}{2} x^2 e^\xi, \quad (3.34)$$

missä ξ on lukujen 0 ja x väliltä. Tällöin

$$0 \leq 1+x \leq 1+x + \frac{1}{2} x^2 e^\xi = e^x \quad (3.35)$$

jokaisella $x \geq -1$. Koska m on positiivinen reaaliluku, niin väite seuraa epäyhtälöketjusta (3.35). \square

Tarvitaan vielä toinen apulause, jonka todistuksessa tarvitaan edellistä apulausetta.

Apulause 3.10. *Olkoon s ja t positiivisia reaalilukuja ja $\{a_n\}$ lukujono, joka toteuttaa seuraavat ehdot:*

$$a_0 \geq -\frac{t}{s} \quad \text{ja} \quad a_{n+1} \leq (1+s)a_n + t. \quad (3.36)$$

Tällöin

$$a_{n+1} \leq e^{(n+1)s} \left(\frac{t}{s} + a_0 \right) - \frac{t}{s}. \quad (3.37)$$

Todistus. (Vrt. [5, s. 552–553]) Osoitetaan ensin induktiolla, että

$$a_{n+1} \leq (1+s)^{n+1} a_0 + t \sum_{j=1}^{n+1} (1+s)^{j-1}, \quad (3.38)$$

jokaisella $n = 0, 1, 2, \dots$. Oletuksen (3.36) mukaan

$$a_1 \leq (1+s)a_0 + t, \quad (3.39)$$

joten väite (3.38) on tosi, kun $n = 0$. Oletetaan sitten, että väite (3.38) on tosi, kun $n = k$, eli

$$a_{k+1} \leq (1+s)^{k+1} a_0 + t \sum_{j=1}^{k+1} (1+s)^{j-1}. \quad (3.40)$$

Osoitetaan, että tällöin väite (3.38) on myös tosi, kun $n = k + 1$. Oletuksen (3.36) mukaan

$$\begin{aligned}
 a_{k+2} &\leq (1+s)a_{k+1} + t \\
 &\stackrel{(3.40)}{\leq} (1+s) \left((1+s)^{k+1} a_0 + t \sum_{j=1}^{k+1} (1+s)^{j-1} \right) + t \\
 &= (1+s)^{k+2} a_0 + t \sum_{j=1}^{k+1} (1+s)^j + t \\
 &= (1+s)^{k+2} a_0 + t \sum_{j=1}^{k+2} (1+s)^{j-1},
 \end{aligned} \tag{3.41}$$

joten väite on myös tosi, kun $n = k + 1$. Induktioperiaatteen nojalla epäyhtälö (3.38) on tosi jokaisella $n = 0, 1, 2, \dots$

Nyt voidaan käyttää hyödyksi geometrisen sarjan osasummaa

$$S_k = \sum_{j=1}^k ar^{j-1} = \frac{a(1-r^k)}{1-r}, \tag{3.42}$$

missä a on sarjan ensimmäinen termi ja $r \neq 1$ [1, s. 481]. Epäyhtälö (3.38) saadaan kaavan (3.42) avulla muotoon

$$\begin{aligned}
 a_{n+1} &\leq (1+s)^{n+1} a_0 + \frac{t}{s} [(1+s)^{n+1} - 1] \\
 &= (1+s)^{n+1} \left(a_0 + \frac{t}{s} \right) - \frac{t}{s}
 \end{aligned} \tag{3.43}$$

Apulauseen 3.9 mukaan

$$0 \leq (1+s)^{n+1} \leq e^{(n+1)s}, \tag{3.44}$$

eli

$$a_{n+1} \leq e^{(n+1)s} \left(a_0 + \frac{t}{s} \right) - \frac{t}{s}. \tag{3.45}$$

□

Nyt voidaan todistaa lause, joka antaa ylärajan kokonaisvirheen itseisarvolle.

Lause 3.11. *Olkoon $y(x)$ alkuarvotehtävän*

$$\begin{aligned}
 y'(x) &= f(x, y(x)), \\
 y(a) &= y_0,
 \end{aligned} \tag{3.46}$$

yksikäsitteinen ratkaisu välillä $a \leq x \leq b$. Olkoon y_1, y_2, \dots, y_N Eulerin menetelmällä (3.14) saadut alkuarvotehtävän (3.46) numeeriset ratkaisut. Jos funktio f toteuttaa Lipschitz-ehdon muuttujan y suhteen joukossa $D = \{(x, y) \mid a \leq x \leq b, y \in \mathbb{R}\}$, ja on olemassa sellainen reaalityö M , että

$$\max_{x \in [a, b]} |y''(x)| \leq M, \tag{3.47}$$

niin

$$|e_n| \leq \frac{hM}{2L} \left(e^{L(x_n-a)} - 1 \right). \quad (3.48)$$

Todistus. (Vrt. [5, s. 553]) Ensinnäkin nähdään, että kaava (3.48) on voimassa, kun $n = 0$. Tällöin $e_0 = y(x_0) - y_0 = y_0 - y_0 = 0$, ja sijoittamalla kaavaan (3.48) $n = 0$, saadaan

$$|e_0| \leq \frac{hM}{2L} \left(e^{L(a-a)} - 1 \right) = 0. \quad (3.49)$$

Osoitetaan vielä, että kaava (3.48) on voimassa, kun $n = 1, 2, 3, \dots, N$.

Taylorin kaavan avulla saadaan

$$y(x_{n+1}) = y(x_n) + hf(x_n, y(x_n)) + \frac{h^2}{2}y''(\xi_n), \quad (3.50)$$

kun $n = 0, 1, 2, \dots, N - 1$. Eulerin menetelmällä taas saadaan

$$y_{n+1} = y_n + hf(x_n, y_n), \quad (3.51)$$

kun $n = 0, 1, 2, \dots, N - 1$. Vähentämällä yhtälö (3.51) yhtälöstä (3.50) puolittain, saadaan

$$y(x_{n+1}) - y_{n+1} = y(x_n) - y_n + h[f(x_n, y(x_n)) - f(x_n, y_n)] + \frac{h^2}{2}y''(\xi_n). \quad (3.52)$$

Otetaan yhtälöstä (3.52) itseisarvo puolittain, eli

$$\begin{aligned} |y(x_{n+1}) - y_{n+1}| &= \left| y(x_n) - y_n + h[f(x_n, y(x_n)) - f(x_n, y_n)] + \frac{h^2}{2}y''(\xi_n) \right| \\ &\leq |y(x_n) - y_n| + h|f(x_n, y(x_n)) - f(x_n, y_n)| + \frac{h^2}{2}|y''(\xi_n)|. \end{aligned} \quad (3.53)$$

Nyt voidaan käyttää hyödyksi funktion f Lipschitz-jatkuvuutta ja alkuarvotteävän ratkaisun toisen derivaatan $y''(x)$ itseisarvon ylärajaa. Tällöin

$$\begin{aligned} |y(x_{n+1}) - y_{n+1}| &\leq |y(x_n) - y_n| + hL|y(x_n) - y_n| + \frac{h^2M}{2} \\ &= (1 + hL)|y(x_n) - y_n| + \frac{h^2M}{2}. \end{aligned} \quad (3.54)$$

Huomataan, että (3.54) toteuttaa Apulauseen 3.10 oletukset, kun $s = hL$, $t = h^2M/2$ ja $a_n = |y(x_n) - y_n|$. Nyt siis Apulauseen 3.10 mukaan

$$|y(x_{n+1}) - y_{n+1}| \leq e^{(n+1)hL} \left(\frac{hM}{2L} - |y(x_0) - y_0| \right) - \frac{hM}{2L} \quad (3.55)$$

Todistuksen alussa todettiin, että $y(x_0) - y_0 = 0$, joten

$$|y(x_{n+1}) - y_{n+1}| \leq \frac{hM}{2L} \left(e^{(n+1)hL} - 1 \right). \quad (3.56)$$

Tiedetään, että $x_{n+1} = x_0 + (n+1)h = a + (n+1)h$, joten

$$(n+1)h = x_{n+1} - a. \quad (3.57)$$

Sijoittamalla edellinen epäyhtälöön (3.56), saadaan

$$|y(x_{n+1}) - y_{n+1}| \leq \frac{hM}{2L} \left(e^{(x_{n+1}-a)L} - 1 \right). \quad (3.58)$$

□

Lauseen 3.11 avulla saadaan kaksi keskeistä tulosta. Ensinnäkin Eulerin menetelmän kokonaisvirhe

$$e_n = O(h). \quad (3.59)$$

Toiseksi

$$\lim_{h \rightarrow 0} \frac{hM}{2L} \left(e^{(x_n-a)L} - 1 \right) = 0 \quad (3.60)$$

jokaisella $n = 0, 1, 2, \dots, N$, joten $|e_n| \rightarrow 0$, kun $h \rightarrow 0$. Määritelmän 3.5 mukaan Eulerin menetelmä on siis suppeneva. Huomioitavaa on se, että vakiota M ei tunneta. Jos kuitenkin funktion f osittaisderivaatat f_y ja f_x ovat olemassa tarkasteltavalla välillä, niin vakio M voidaan määrittää, sillä

$$y'' = f' = f_x + f_y y' = f_x + f_y f. \quad (3.61)$$

Lisäksi Lauseen 3.11 antama arvio kokonaisvirheen itseisarvion ylärajalle on useille alkuarvoteh-
täville paljon suurempi, mitä kokonaisvirheen itseisarvo todellisuudessa on. [5, s. 552]

4. RUNGE–KUTTA-MENETELMÄT

Tässä luvussa keskitytään Runge–Kutta-menetelmiin. Luvussa 3 esitetyissä menetelmissä on muutamia heikkouksia, jotka liittyvät numeerisen ratkaisun tarkkuuteen ja menetelmien tehokkuuteen. Käytännön ongelmissa halutaan usein mahdollisimman tarkka ratkaisu. Eulerin menetelmä on 1. kertaluvun menetelmä, joten usein tarkan ratkaisun saaminen vaatii pienen askelpituuden h , mikä taas kasvattaa vaadittavien askelten määrää N [2, s. 73]. Taylorin menetelmien kertaluku pystytään saamaan korkeammaksi kuin 1, joten usein niillä saadaan tarkempia ratkaisuja kuin 1. kertaluvun menetelmillä. Taylorin menetelmien käyttäminen vaatii kuitenkin funktion $f(x, y)$ kokonaisderivaattojen laskemista, mikä voi käytännössä osoittautua haastavaksi [2, s. 74]. Tämä johtaa Runge–Kutta-menetelmiin, jotka vaativat vain funktion $f(x, y)$ arvojen laskemista eri pisteissä.

4.1 Eksplisiittiset ja implisiittiset Runge–Kutta-menetelmät

Yksi tapa johtaa Eulerin menetelmä oli yhtälössä

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x)) dx. \quad (4.1)$$

olevan integraalin approksimointi. Vastaavalla tavalla saadaan myös useat Runge–Kutta-menetelmät [9, s. 38]. Yleinen m -vaiheinen Runge–Kutta-menetelmä on muotoa

$$y_{n+1} = y_n + h \sum_{i=1}^m b_i k_i, \quad (4.2)$$

$$k_i = f \left(x_n + c_i h, y_n + h \sum_{j=1}^m a_{i,j} k_j \right), \quad i = 1, 2, \dots, m.$$

Runge–Kutta-menetelmässä esiintyvät kertoimet annetaan tyypillisesti taulukossa, jota kutsutaan Butcherin taulukoksi. Yleistä m -vaiheista Runge–Kutta-menetelmää kuvaava Butcherin taulukko on muotoa

$$\begin{array}{c|cccc} c_1 & a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ c_2 & a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_m & a_{m,1} & a_{m,2} & \cdots & a_{m,m} \\ \hline & b_1 & b_2 & \cdots & b_m \end{array}, \quad (4.3)$$

mikä voidaan kirjoittaa myös lyhyemmin muodossa

$$\frac{\mathbf{c} \mid A}{\mid \mathbf{b}^T}. \quad (4.4)$$

Jos $a_{i,j} \neq 0$, jollain $j > i$, niin Runge–Kutta-menetelmää sanotaan implisiittiseksi Runge–Kutta-menetelmäksi. [10, s. 149–150] Menetelmien kertoimia $b_i, i = 1, 2, \dots, m$ kutsutaan painokertoimiksi ja kertoimia $c_i, i = 1, 2, \dots, m$ kutsutaan solmuiksi [9, s. 38]. Nämä kertoimet ja matriisi A pyritään valitsemaan siten, että menetelmän virhe saadaan mahdollisimman pieneksi. Lisäksi vaaditaan, että ehto

$$c_i = \sum_{j=1}^m a_{i,j}, \quad (4.5)$$

on voimassa kun $i = 1, 2, \dots, m$. [2, s. 80]

Jos $a_{i,j} = 0$, kun $j \geq i, i = 1, 2, \dots, m$, niin jokainen funktio k_i saadaan määritettyä aiemmin määritettyjen funktioiden $k_l, l = 1, 2, \dots, i-1$ avulla. Tällaista Runge–Kutta-menetelmää sanotaan eksplisiittiseksi Runge–Kutta-menetelmäksi. [10, s. 150] Eksplisiittisen menetelmän Butcherin taulukosta jätetään usein merkittömättä matriisin A diagonaalilla ja sen yläpuolella olevat nollat [9, s. 38], eli eksplisiittisen menetelmän Butcherin taulukko on muotoa

$$\begin{array}{c|cccc} 0 & & & & \\ c_2 & a_{2,1} & & & \\ c_3 & a_{3,1} & a_{3,2} & & \\ \vdots & \vdots & \vdots & \ddots & \\ c_m & a_{m,1} & a_{m,2} & \cdots & a_{m,m-1} \\ \hline & b_1 & b_2 & \cdots & b_{m-1} & b_m \end{array} \quad (4.6)$$

Eksplisiittinen m -vaiheinen Runge–Kutta-menetelmä on siis

$$\begin{aligned} y_{n+1} &= y_n + h \sum_{i=1}^m b_i k_i, \\ k_1 &= f(x_n, y_n), \\ k_i &= f\left(x_n + c_i h, y_n + h \sum_{j=1}^{i-1} a_{i,j} k_j\right), \quad i = 2, \dots, m. \end{aligned} \quad (4.7)$$

Menetelmän (4.7) jokaisella askeleella lasketaan siis funktion f arvo m kertaa. [7, s. 341] Tässä kohtaa on myös huomata, että Luvussa 3 esitetty Eulerin menetelmä on eksplisiittinen 1-vaiheinen Runge–Kutta-menetelmä, jonka Butcherin taulukko on

$$\frac{0 \mid}{\mid 1}. \quad (4.8)$$

Huomataan, että menetelmät (4.2) ja (4.7) ovat muotoa (3.5), eli ne ovat eksplisiittisiä yksiaskelmenetelmiä.

Huomautus 4.1. Tässä kohtaa ei tule sekoittaa keskenään Runge–Kutta-menetelmän eksplisiittisyyttä/implisiittisyyttä ja yksiaskelmenetelmän eksplisiittisyyttä/implisiittisyyttä. Runge–Kutta-menetelmän eksplisiittisyys/implisiittisyys riippuu matriisista A . Yksiaskelmenetelmän eksplisiittisyys/implisiittisyys taas riippuu siitä, onko lisäysfunktio riippuvainen muuttujasta y_{n+1} . Sekä implisiittiset että eksplisiittiset Runge–Kutta-menetelmät ovat eksplisiittisiä yksiaskelmenetelmiä.

Tarkastellaan vielä, milloin Runge–Kutta-menetelmä on konsistenssi. Runge–Kutta-menetelmän lisäysfunktio

$$\Phi(x_n, y_n, h) = \sum_{i=1}^m b_i k_i, \quad (4.9)$$

missä funktiot k_i ovat muotoa (4.2) tai (4.7).

Lause 4.2. *Runge–Kutta-menetelmä on konsistenssi, jos*

$$\sum_{i=1}^m b_i = 1. \quad (4.10)$$

Todistus. Kun $h = 0$, niin

$$\Phi(x_n, y(x_n), 0) = \sum_{i=1}^m b_i f(x_n, y(x_n)) = f(x_n, y(x_n)) \sum_{i=1}^m b_i. \quad (4.11)$$

Oletuksen nojalla

$$\Phi(x_n, y(x_n), 0) = f(x_n, y(x_n)), \quad (4.12)$$

eli ehto (3.13) on voimassa. □

Jos ehto (4.10) on voimassa, niin lisäysfunktio (4.9) on painotettu keskiarvo funktioista k_i . Huomioarvoista on, että funktiot k_i kuvaavat alkuarvotettävän ratkaisun y derivaatan arvoja eri pisteissä [10, s. 151], eli lisäysfunktio kuvaa alkuarvotettävän ratkaisun keskimääräistä suuntaa eri pisteissä.

Tarkastellaan tästä eteenpäin eksplisiittisiä Runge–Kutta-menetelmiä, mutta jätetään eksplisiittisyys mainitsematta.

4.2 2-vaiheinen Runge–Kutta-menetelmä

Tässä alaluvussa etsitään 2-vaiheisessa Runge–Kutta-menetelmässä esiintyvät kertoimet. Tällaisen menetelmän Butcherin taulukko on

$$\begin{array}{c|cc} 0 & & \\ c_2 & a_{2,1} & \\ \hline & b_1 & b_2 \end{array}. \quad (4.13)$$

Ehdon (4.5) perusteella $a_{2,1} = c_2$, joten 2-vaiheinen Runge–Kutta-menetelmä on

$$\begin{aligned} y_{n+1} &= y_n + h(b_1 k_1 + b_2 k_2), \\ k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + c_2 h, y_n + c_2 h k_1), \end{aligned} \quad (4.14)$$

minkä lisäysfunktio $\Phi(x_n, y_n, h) = b_1 k_1 + b_2 k_2$.

Menetelmässä (4.14) esiintyvät kertoimet b_1, b_2 ja c_2 voidaan määrittää Taylorin kaavan avulla. Taylorin kaavalla saadaan

$$y(x_{n+1}) = y(x_n) + y'(x_n)h + \frac{y''(x_n)}{2}h^2 + \frac{y'''(x_n)}{6}h^3 + \mathcal{O}(h^4). \quad (4.15)$$

Oletetaan, että funktio f on tarpeeksi sileä, jolloin derivaatat saadaan määritettyä differentiaaliyh-tälön avulla. Koska $y' = f(x, y)$, niin

$$y'' = f'(x, y) = f_x + f_y y' = f_x + f f_y \quad (4.16)$$

ja

$$\begin{aligned} y''' &= f''(x, y) = (f_x + f_y f)' \\ &= f_{xx} + f_{xy} y' + (f_{yx} + f_{yy} y') f + f_y f' \\ &= f_{xx} + f_{xy} f + (f_{yx} + f_{yy} f) f + f_y (f_x + f f_y) \\ &= f_{xx} + 2f f_{xy} + f^2 f_{yy} + f_y (f_x + f f_y). \end{aligned} \quad (4.17)$$

Tarkastelun helpottamiseksi otetaan käyttöön merkinnät

$$s = f_x + f f_y, \quad t = f_{xx} + 2f f_{xy} + f^2 f_{yy}. \quad (4.18)$$

Näillä merkinnöillä yhtälö (4.15) saadaan muotoon

$$y(x_{n+1}) = y(x_n) + hf + \frac{1}{2}h^2 s + \frac{1}{6}h^3 (t + f_y s) + \mathcal{O}(h^4), \quad (4.19)$$

missä funktio f ja sen osittaisderivaatat ovat määritetty pisteessä $(x_n, y(x_n))$. [10, s. 153–154] Menetelmän (4.14) lisäysfunktio täytyy saada myös muotoon, jossa esiintyy funktio f ja sen osittaisderivaattoja. Tämän jälkeen voidaan löytää menetelmän kertoimet.

Kahden muuttujan funktion Taylorin sarjan avulla funktio k_2 voidaan kirjoittaa muodossa

$$k_2 = f + hc_2(f_x + f f_y) + \frac{1}{2}h^2 c_2^2 (f_{xx} + f f_{xy} + f^2 f_{yy}) + \mathcal{O}(h^3), \quad (4.20)$$

ja edelleen merkintöjen (4.18) avulla muodossa

$$k_2 = f + hc_2 s + \frac{1}{2}h^2 c_2^2 t + \mathcal{O}(h^3), \quad (4.21)$$

missä funktio f ja sen osittaisderivaatat ovat määritetty pisteessä (x_n, y_n) [10, s. 154]. Tällöin

menetelmän (4.14) lisäysfunktio, saadaan muotoon

$$\begin{aligned}\Phi(x_n, y_n, h) &= b_1 k_1 + b_2 k_2 \\ &= (b_1 + b_2)f + hb_2 c_2 s + \frac{1}{2}h^2 b_2 c_2^2 t + O(h^3).\end{aligned}\quad (4.22)$$

Nyt voidaan laskea paikallinen diskreointivirhe yhtälön (4.19) ja lisäysfunktion (4.22) avulla. Näiden avulla saadaan

$$\begin{aligned}\tau_n &= (1 - b_1 - b_2)f + \left(\frac{1}{2}s - b_2 c_2 s\right)h + \left(\frac{1}{6}(t + f_y s) - \frac{1}{2}b_2 c_2^2 t\right)h^2 + O(h^3) \\ &= (1 - b_1 - b_2)f + \left(\frac{1}{2} - b_2 c_2\right)sh + \left[\left(\frac{1}{6} - \frac{1}{2}b_2 c_2^2\right)t + \frac{1}{6}f_y s\right]h^2 + O(h^3),\end{aligned}\quad (4.23)$$

missä funktio f ja sen osittaisderivaatat ovat määritetty pisteessä $(x_n, y(x_n))$. [11, s. 327]

Paikallinen diskreointivirhe halutaan mahdollisimman lähelle nollaa [3, s. 422]. Yhtälöstä (4.23) nähdään, että funktion f ja askelpituuden h kertoimet ovat nollia, kun

$$b_1 + b_2 = 1, \quad b_2 c_2 = \frac{1}{2}.\quad (4.24)$$

Tällöin paikallinen diskreointivirhe

$$\tau_n = \left[\left(\frac{1}{6} - \frac{1}{4}c_2\right)t + \frac{1}{6}f_y s\right]h^2 + O(h^3)\quad (4.25)$$

Huomataan, että ei ole sellaista kertoimen c_2 arvoa, jolla askelpituuden neliön h^2 kerroin on nolla jokaisella funktiolla f . Esimerkiksi tarkasteltaessa alkuarvottehtävää $y' = y$, $y(0) = 1$, paikallinen diskreointivirhe

$$\tau_n = \frac{1}{6}h^2 y(x_n) + O(h^3) = \frac{1}{6}h^2 e^{x_n} + O(h^3).\quad (4.26)$$

Paikallisen diskreointivirheen (4.25) perusteella 2-vaiheisen Runge–Kutta-menetelmän kertaluku on siis 2, kun ehdot (4.24) ovat voimassa. [11, s. 327]

Ratkaistaan vielä yhtälöryhmästä (4.24) kertoimet b_1 ja b_2 kertoimen c_2 avulla. Ratkaisuksi saadaan

$$b_1 = 1 - \frac{1}{2c_2}, \quad b_2 = \frac{1}{2c_2}, \quad c_2 \neq 0,\quad (4.27)$$

eli 2. kertaluvun 2-vaiheisia Runge–Kutta-menetelmiä on ääretön määrä [10, s. 154–155]. Lisäksi huomataan, että Lauseen 4.2 mukaan nämä menetelmät ovat konsistensseja.

2-vaiheisen Runge–Kutta-menetelmän, jonka kertaluku on 2, Butcherin taulukko on siis

$$\begin{array}{c|cc} 0 & & \\ c_2 & c_2 & \\ \hline & 1 - \frac{1}{2c_2} & \frac{1}{2c_2} \end{array},\quad (4.28)$$

missä $c_2 \neq 0$.

Esimerkki 4.3. Valitaan $c_2 = \frac{2}{3}$. Tällöin funktion t kerroin on nolla paikallisessa diskreetointivirheessä (4.25). Tämän menetelmän Butcherin taulukko on

$$\begin{array}{c|cc} 0 & & \\ \frac{2}{3} & \frac{2}{3} & \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}, \quad (4.29)$$

eli menetelmä on

$$\begin{aligned} y_{n+1} &= y_n + \frac{h}{4}(k_1 + 3k_2), \\ k_1 &= f(x_n, y_n), \\ k_2 &= f\left(x_n + \frac{2}{3}h, y_n + \frac{2}{3}hk_1\right). \end{aligned} \quad (4.30)$$

Tätä menetelmää kutsutaan optimaaliseksi RK2-menetelmäksi tai Ralsonin menetelmäksi [5, s. 572].

Vastaavalla tavalla voidaan löytää ehdot 3. ja 4. kertaluvun menetelmän kertoimille. 4. kertaluvun tapauksessa Runge–Kutta-menetelmä on 4-vaiheinen. Tällöin menetelmien kertoimien löytämisestä tulee huomattavasti vaikeampaa, sillä tarkastelussa tarvitaan funktiot $k_i, i = 1, 2, 3, 4$. [9, s. 41] Tunnetuin 4. kertaluvun Runge–Kutta-menetelmä on niin sanottu Klassinen Runge–Kutta-menetelmä, jonka Butcherin taulukko on

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}. \quad (4.31)$$

Jos sanotaan, että alkuarvotehtävä on ratkaistu Runge–Kutta-menetelmällä, niin yleensä tehtävä on ratkaistu Klassisella Runge–Kutta-menetelmällä. [10, s. 156]

Vielä suuremman kertaluvun menetelmien johtamiseen tarvitaan graafiteoriaan perustuvia keinoja. Lisäksi jos halutaan kertaluvun $p \geq 5$ menetelmä, niin tarvitaan m -vaiheinen Runge–Kutta-menetelmä, missä $m > p$. Esimerkiksi 5. kertaluvun Runge–Kutta-menetelmä vaatii kuusi vaihetta, eli $p = 5$ ja $m = 6$. [9, s. 41]

5. RUNGE–KUTTA-MENETELMÄT KÄYTÄNNÖSSÄ

Tässä luvussa ratkaistaan alkuarvotettava, ja vertaillaan eri menetelmien kokonaisvirheitä. Alkuarvotettävän ratkaisemiseen ja menetelmien vertailuun käytetään Liitteiden A ja B MATLAB-koodeja.

5.1 Alkuarvotettävän numeerinen ratkaisu: putoava pallo

Tarkastellaan maanpinnan läheisyydessä putoavaa palloa. Putoavaan palloon kohdistuu kaksi voimaa: maan painovoima G ja ilmanvastus F_D . Ilmanvastus riippuu kappaleen muodosta, väliaineen ominaisuuksista ja kappaleen vauhdista. Ilmanvastus on likimain suoraan verrannollinen kappaleen vauhtiin, kun kappaleen vauhti on pieni. Suuremmilla nopeuksilla ilmanvastus on taas likimain suoraan verrannollinen kappaleen vauhdin neliöön. [12, s. 134] Kuvassa 5.1 on esitetty pallon voimakuvio, kun pallo on kiihtyvässä liikkeessä kuvan y -askelin suuntaan.

Oletetaan, että ilmanvastuksen suuruus on suoraan verrannollinen pallon vauhdin neliöön v^2 , eli

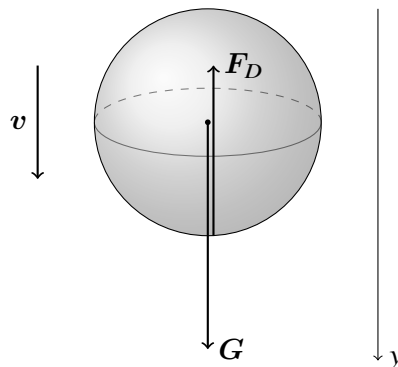
$$F_D = \|F_D\| = kv^2, \quad k > 0 \quad (5.1)$$

missä k on verrannollisuuskerroin. Lisäksi painovoiman suuruus

$$G = \|G\| = mg, \quad (5.2)$$

missä m on pallon massa ja g on putoamiskiihtyvyys maanpinnan lähellä. Nyt siis Newtonin 2. lain mukaan

$$mg - kv(t)^2 = ma(t), \quad (5.3)$$



Kuva 5.1. Putoavaan palloon kohdistuvat voimat.

eli pallon kiihtyvyys

$$a(t) = g - \frac{k}{m}v(t)^2. \quad (5.4)$$

Jos pallo on pudotuksen alussa paikallaan, eli $v(0) = 0$, niin pallo on kiihtyvässä liikkeessä Kuvan 5.1 y -akselin suuntaan. Tällöin nopeus y -akselin suuntaan kasvaa, joten yhtälön (5.4) mukaan kiihtyvyys pienenee. Jossain vaiheessa yhtälön (5.4) vauhtitermi tulee siis yhtäsuureksi kuin putoamiskiihtyvyys g , eli kiihtyvyys on nolla, ja pallo saavuttaa maksiminopeutensa. Olettaen, että pallo ei kerkeä osua maahan ennen tätä. Maksiminopeus v_{\max} saadaan ratkaistua yhtälöstä (5.4), kun kiihtyvyys asetetaan nolaksi, eli

$$v_{\max} = \pm \sqrt{\frac{mg}{k}}. \quad (5.5)$$

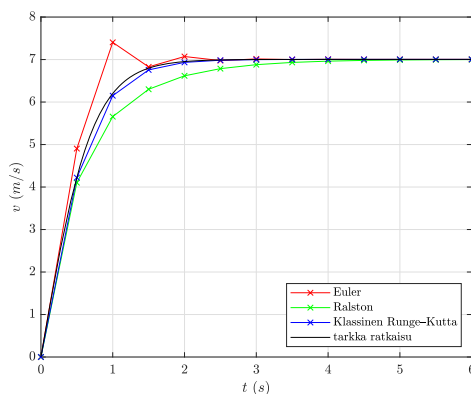
Hyväksytään näistä vain positiivinen ratkaisu, sillä negatiivinen ratkaisu ei ole fyysikaalisesti järkevä. (Vrt. [12, s. 135]) Saatiin siis ratkaistua putoavan pallon maksiminopeus. Ei kuitenkaan vielä tiedetä pallon nopeutta, jollain mielivaltaisella hetkellä $t > 0$ s. Lähdetään ratkaisemaan tätä ongelmaa differentiaaliyhtälön avulla.

Kiihtyvyys määritellään nopeuden derivaattana ajan suhteen [12, s. 25], joten yhtälöstä (5.4) saadaan 1. kertaluvun differentiaaliyhtälö. Jos myös vaaditaan, että pallo on hetkellä $t = 0$ s paikallaan, niin saadaan alkuarvotehtävä

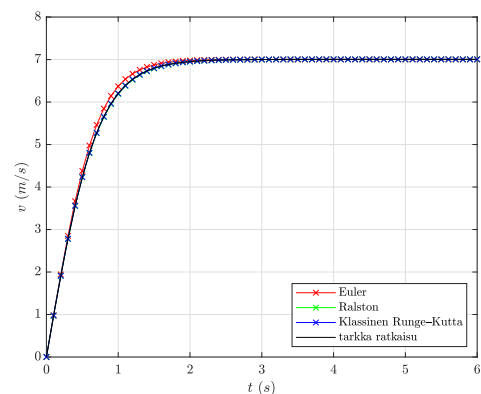
$$v'(t) = g - \frac{k}{m}v(t)^2, \quad v(0) = 0. \quad (5.6)$$

Tämä alkuarvotehtävä voitaisiin ratkaista esimerkiksi separoimalla, mutta ratkaistaan se numeerisesti Eulerin menetelmällä, Ralstonin menetelmällä ja Klassisella Runge–Kutta menetelmällä aikavälillä $[0, 6]$ (s). Ratkaisemiseen käytetään Liitteen A MATLAB-koodeja.

Nyt ei olla kiinnostuneita verrannollisuuskertoimen k todellisesta arvosta, vaan alkuarvotehtävän ratkaisemisesta, joten yksinkertaisuuden vuoksi valitaan verrannollisuuskertoimeksi $k = 1$ kg/m. Putoamiskiihtyvyys $g = 9.81$ m/s² maan pinnan lähellä, ja lisäksi valitaan pallon massaksi $m = 5$ kg. Tällöin Liitteen B.2 MATLAB-koodilla saadaan Kuvan 5.2 mukaiset numeeriset ratkaisut.



(a) Askelpituus $h = 0.5$, eli $N = 12$.



(b) Askelpituus $h = 0.1$, eli $N = 60$.

Kuva 5.2. Alkuarvotehtävän (5.6) numeeriset ratkaisut askelpituuksilla $h = 0.5$ ja $h = 0.1$.

Kuvassa 5.2 on esitetty myös alkuarvottehtävän (5.6) tarkka ratkaisu

$$v(t) = \sqrt{\frac{mg}{k}} \tanh\left(\sqrt{\frac{gk}{m}} t\right), \quad (5.7)$$

mikä saatiin MATLAB:n symbolisen laskennan toolboxin `dsolve`-komennolla.

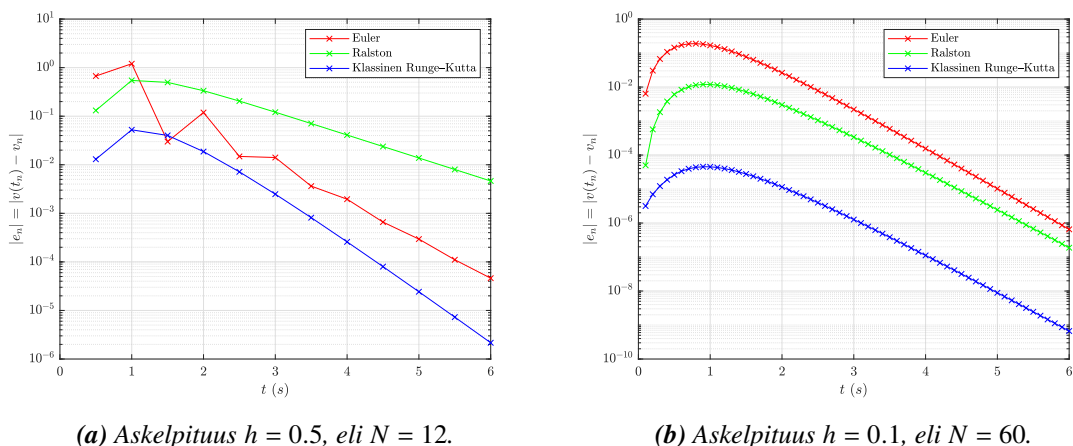
Kuvasta 5.2a nähdään, että Eulerin menetelmän ja Ralstonin menetelmän antamat ratkaisut eroavat selvästi tarkasta ratkaisusta, mutta molemmat kuitenkin lähestyvät tarkkaa ratkaisua, kun tarkka ratkaisu tasoittuu. Nähdään myös, että askelpituudella $h = 0,5$ Klassinen Runge–Kutta-menetelmä antaa jo suhteellisen hyvän ratkaisun. Kun askelpituutta pienennetään, niin Kuvasta 5.2b nähdään selvästi, että Eulerin menetelmän ja Ralstonin menetelmän antamat ratkaisut lähestyvät tarkkaa ratkaisua. Huomataan myös, että askelpituudella $h = 0,5$ Eulerin menetelmän antama ratkaisun suurin virhe on kohdassa $t = 1$ s, jonka läheisyydessä tarkan ratkaisun derivaattaa muuttuu nopeasti. Tällaiset nopeat muutokset vaativat pienen askelpituuden, jos halutaan saada hyvä ratkaisu. Tämä ongelma voidaan korjata käyttämällä sellaista numeerista menetelmää, missä askelpituus ei ole vakio ratkaisun edetessä, ks. [5, s. 609–613].

5.2 Menetelmien vertailu

Tarkastellaan edelleen alkuarvottehtävää (5.4) samoilla parametreilla m ja k , ja vertaillaan eri menetelmien tekemiä virheitä. Lasketaan ensin kokonaisvirheiden itseisarvot

$$|e_n| = |v(t_n) - v_n|, \quad n = 1, 2, \dots, N, \quad (5.8)$$

missä $v(t_n)$ on alkuarvottehtävän tarkka ratkaisu pisteessä t_n ja v_n on numeerisella menetelmällä saatu ratkaisu pisteessä t_n . Käytetään askelpituuksia $h = 0,5$ ja $h = 0,1$. Kokonaisvirheiden itseisarvot on esitetty Kuvassa 5.3, joka saatiin Liitteen B.2 MATLAB-koodilla.



Kuva 5.3. Kokonaisvirheen itseisarvo $|e_n| = |v(t_n) - v_n|$, kun $n = 1, 2, \dots, N$.

Huomataan, että jokaisen menetelmän kokonaisvirheen itseisarvon maksimiarvo on pienempi askelpituudella $h = 0,1$ kuin askelpituudella $h = 0,5$, eli tällöin saadaan tarkempia ratkaisuja, kuten

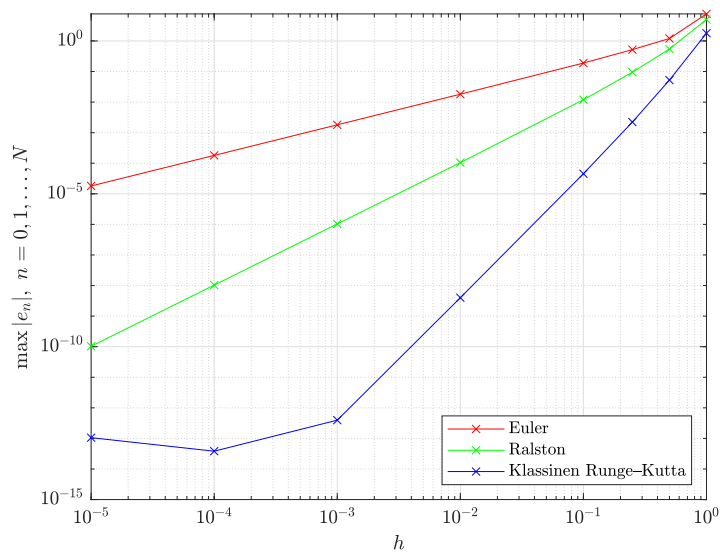
Taulukko 5.1. Käytetyt askelpituudet h ja niitä vastaavat askelten määrät N .

| h | 1 | 0,5 | 0,25 | 0,1 | 0,01 | 0,001 | 0,0001 | 0,00001 |
|-----------|---|-----|------|-----|------|-------|--------|---------|
| $N = 6/h$ | 6 | 12 | 24 | 60 | 600 | 6000 | 60000 | 600000 |

Kuvasta 5.2 huomattiin. Kuvasta 5.2b ei ollut mahdollista erottaa Ralstonin menetelmän ja Klassisen Runge–Kutta-menetelmän antamia ratkaisuja toisistaan. Nyt kuitenkin Kuvasta 5.3b nähdään, että Klassisen Runge–Kutta-menetelmän kokonaisvirheiden itseisarvot ovat selvästi pienempiä kuin Ralstonin menetelmän kokonaisvirheiden itseisarvot. Nähdään myös, että suurimmat kokonaisvirheet tulevat sillä välillä, missä tarkan ratkaisun derivaatta muuttuu nopeasti. Ja pienimmät kokonaisvirheet tulevat silloin, kun tarkka ratkaisu tasoittuu.

Lasketaan vielä eri menetelmien maksimi kokonaisvirheitä eri askelpituuksilla. Käytetään samoja parametrien arvoja m ja k kuin aiemmin, ja tarkastellaan edelleen väliä $t \in [0, 6]$. Ratkaistaan siis alkuarvottehtävä (5.6) käyttäen Taulukon 5.1 askelpituuksia. Liitteen B.3 MATLAB-koodilla saadaan Kuva 5.4, jossa on esitetty eri menetelmien kokonaisvirheen itseisarvon maksimi eri askelpituuksilla. Kuvasta nähdään, että klassinen Runge–Kutta-menetelmä lähestyy tarkkaa ratkaisua nopeiten, kun askelpituutta pienennetään. Eulerin menetelmä taas lähestyy tarkkaa ratkaisua hitaimmin, mikä oli odotettavissa, sillä Eulerin menetelmän kertaluku on vain 1.

Alaluvussa 3.4 osoitettiin, että Eulerin menetelmän kokonaisvirhe $e_n = O(h)$. Kuvan 5.4 tulokset tukevat tätä. Lisäksi kuvan mukaan Ralstonin menetelmän kokonaisvirhe näyttäisi olevan $O(h^2)$, ja Klassisen Runge–Kutta-menetelmän kokonaisvirhe näyttäisi olevan $O(h^4)$. Molemmat näistä väittämistä on totta, ja yleisesti jos eksplisiittisen yksiaskelmenetelmän kertaluku on p , eli $\tau_n = O(h^p)$, niin myös kokonaisvirhe $e_n = O(h^p)$ [5, s. 600–601]. Kuvassa 5.4 huomioitavaa on myös se, että Klassinen Runge–Kutta-menetelmän kokonaisvirheen itseisarvon maksimi alkaa kasvamaan, kun askelpituutta pienennetään tarpeeksi. Tämä johtuu luultavasti siitä, että MATLAB käyttää laskennassa liukulukuja.

**Kuva 5.4.** Kokonaisvirheen itseisarvon maksimi eri askelpituuksilla

Kuvasta 5.4 nähdään, että askelpituudella $h = 0,1$ kokonaisvirheiden itseisarvojen maksimeilla on jo huomattava ero. Menetelmien välinen ero korostuu vielä enemmän, kun käytetään pienempää askelpituutta. Täytyy kuitenkin pohtia, että kannattaako askelpituuden pienentäminen, sillä tällöin askelten määrä kasvaa, kuten Taulukosta 5.1 nähdään.

5.3 Vektorimuotoisen alkuarvotehtävän ratkaiseminen

Alaluvun 5.1 tarkastelusta saatiin ratkaistua putoavan pallon vauhti, mutta pallon paikkaan ei päästy suoraan käsiksi. Nopeus määritellään paikan derivaattana ajan suhteen [12, s. 25], joten yhtälöstä (5.4) saadaan 2. kertaluvun differentiaaliyhtälö

$$x''(t) = g - \frac{k}{m}x'(t)^2, \quad (5.9)$$

missä $x(t)$ on pallon paikka hetkellä t Kuvan 5.1 y-akselin suuntaan. Tätä ei voida suoraan ratkaista työssä esitellyillä numeerisilla menetelmillä, mutta se voidaan muuttaa vektorimuotoon, joka voidaan ratkaista tämän työn menetelmillä, kun ne kirjoitetaan vektorimuodossa.

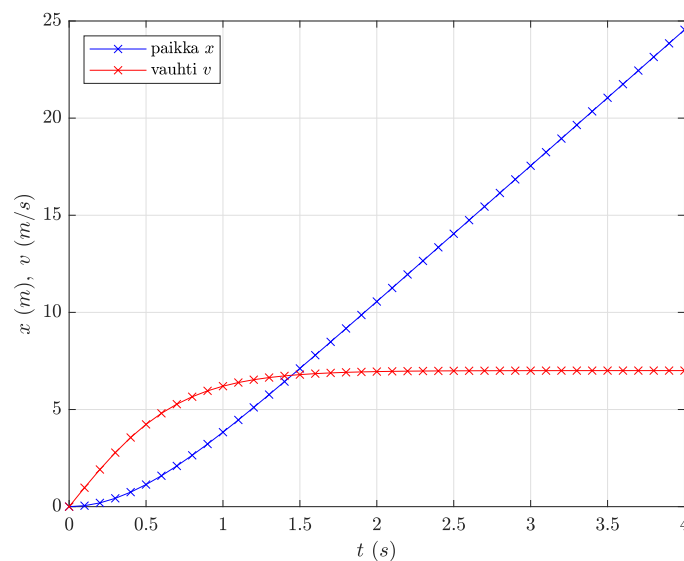
Tiedetään, että $x'(t) = v(t)$, joten

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \quad (5.10)$$

missä

$$\mathbf{y} = \begin{bmatrix} x \\ v \end{bmatrix}, \quad \mathbf{f}(t, \mathbf{y}) = \begin{bmatrix} v \\ g - \frac{k}{m}v^2 \end{bmatrix}. \quad (5.11)$$

Ratkaistaan tämä Klassisella Runge–Kutta-menetelmällä, kun valitaan alkuehdoksi $\mathbf{y}(0) = \mathbf{y}_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, eli $x(0) = 0$ (m) ja $v(0) = 0$ (m/s). Liitteen B.4 MATLAB-koodilla saadaan Kuvan 5.5 mukainen ratkaisu.



Kuva 5.5. Vektorimuotoisen alkuarvotehtävän (5.10) numeeriset ratkaisut Klassisella Runge–Kutta-menetelmällä, kun askelpituus $h = 0,1$.

Kuvasta 5.5 nähdään, että neljän sekunnin aikana pallo putoaa noin 25 metriä. Tässä ratkaisussa tietenkin oletetaan, että pallo ei kerkeä osua maahan ennen tätä. Lisäksi Kuvista 5.2b ja 5.5 havaitaan, että pallon vauhti $v \approx 7,0$ m/s, kun $t \geq 2,0$ s. Ilman differentiaaliyhtälöitä maksiminopeudeksi saadaan

$$v_{\max} = \sqrt{\frac{mg}{k}} = \sqrt{\frac{1 \text{ kg} \cdot 9.81 \text{ m/s}^2}{1 \text{ kg/m}}} \approx 7,00 \text{ m/s}, \quad (5.12)$$

mikä vastaa differentiaaliyhtälöiden avulla saatua maksiminopeutta.

Tässä luvussa saatiin ratkaistua numeerisesti putoavan pallon paikka ja nopeus, kun tunnettiin siihen kohdistuvat voimat. Myös kiihtyvyys saataisiin ratkaistua, jos numeeriset ratkaisut sijoitettaisiin kaavaan (5.4). Vastaavalla tavalla voidaan tutkia myös muiden kappaleiden liikettä, jos tunnetaan niihin kohdistuvat voimat. Putoavan pallon tapauksessa tarkastelu oli yksiulotteista, ja saatiin systeemi, jossa oli kaksi differentiaaliyhtälöä, ks. (5.10)–(5.11). Kaksi- ja kolmeulotteisissa tapauksissa saadaan systeemejä, joissa on enemmän differentiaaliyhtälöitä.

6. YHTEENVETO

Työssä tarkasteltiin alkuarvotehtävien ratkaisemista numeerisesti. Tarkastelu rajattiin eksplisiittisiin yksiaskelmenetelmiin, jotka ovat yksinkertaisimpia numeerisia menetelmiä alkuarvotehtävien ratkaisemiseen. Alkuarvotehtävän ratkaiseminen yksiaskelmenetelmällä perustuu siihen, että tarkasteltava reaalityöväli $[a, b]$ jaetaan pisteisiin x_0, x_1, \dots, x_N , missä $x_0 = a$ ja $x_N = b$. Alkuehdon perusteella tiedetään ratkaisu pisteessä x_0 . Tämän avulla saadaan numeerinen ratkaisu pisteessä x_1 , jonka avulla saadaan edelleen numeerinen ratkaisu pisteessä x_2 . Näin edetään aina pisteeseen x_N asti.

Näytettiin, että Eulerin menetelmä on konsistenssi ja sen kertaluku on 1. Lisäksi löydettiin yläraja Eulerin menetelmän kokonaisvirheen itseisarvolle. Tämän perusteella nähtiin myös, että Eulerin menetelmä suppenee. Työssä esiteltiin myös Runge–Kutta-menetelmät, ja ratkaistiin eksplisiittisen 2-vaiheisen Runge–Kutta-menetelmän kertoimet. Huomattiin, että 2-vaiheisia 2. kertaluvun Runge–Kutta-menetelmiä on ääretön määrä. Yksi esimerkki tällaisesta menetelmästä on Ralstonin menetelmä (4.29). Lisäksi esiteltiin Klassinen Runge–Kutta-menetelmä (4.31), jonka kertaluku on 4.

Ratkaistiin putoavan pallon liikettä kuvaava alkuarvotehtävä. Tämä alkuarvotehtävä saatiin Newtonin 2. lailla, jonka avulla voidaan myös saada muiden kappaleiden liikettä kuvaavia alkuarvotehtäviä. Lisäksi tutkittiin numeerisen menetelmän kokonaisvirheen itseisarvon maksimia. Huomattiin, että korkeamman kertaluvun menetelmät lähestyivät nopeammin tarkkaa ratkaisua, kun askelpituutta pienennettiin.

Luvussa 5 tarkasteltiin sellaista alkuarvotehtävää, jonka ratkaisemiseen soveltui eksplisiittiset yksiaskelmenetelmät. On kuitenkin myös sellaisia alkuarvotehtäviä, joiden ratkaisemiseen soveltuu paremmin monimutkaisemmat numeeriset menetelmät, kuten Alaluvun 3.1 alussa mainitut moniaskelmenetelmät. Lisäksi askelpituus pidettiin tässä työssä vakiona, mutta on olemassa menetelmiä, joissa askelpituutta muutetaan ratkaisemisen edetessä. Tällöin voidaan päästä tarkempiin ratkaisuihin vähemmällä askeleilla.

LÄHTEET

- [1] R. A. Adams. *Calculus: a complete course*. Pearson Education Limited, 2007. 1015 s.
- [2] U. M. Ascher ja L. R. Petzold. *Computer methods for ordinary differential equations and differential-algebraic equations*. Philadelphia (PA): Society for Industrial ja Applied Mathematics, 1998. 314 s.
- [3] K. E. Atkinson. *An introduction to numerical analysis*. New York: Wiley, 1989. 693 s.
- [4] W. E. Boyce ja R. C. DiPrima. *Elementary differential equations and boundary value problems*. New York: Wiley, 1997. 749 s.
- [5] B. Bradie. *A friendly introduction to numerical analysis*. Upper Saddle River: Pearson, 2006. 933 s.
- [6] J. C. Butcher. *Numerical methods for ordinary differential equations*. Chichester, West Sussex, England: John Wiley & Sons, Ltd, 2016. 540 s.
- [7] W. Gautschi. *Numerical Analysis*. Boston, MA: Birkhäuser Boston, 2012. 610 s.
- [8] D. F. Griffiths ja D. J. Higham. *Numerical Methods for Ordinary Differential Equations Initial Value Problems*. London: Springer London, 2010. 273 s.
- [9] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge: University Press, 2008. 459 s.
- [10] J. D. Lambert. *Numerical methods for ordinary differential systems: the initial value problem*. Chichester: Wiley, 1991. 293 s.
- [11] E. Süli ja D. F. Mayers. *An Introduction to Numerical Analysis*. Cambridge: University Press, 2003. 433 s.
- [12] P. A. Tipler ja G. Mosca. *Physics for scientists and engineers*. New York: W.H. Freeman, 2003. 1356 s.

LIITE A: MATLAB-FUNKTIOT NUMEERISILLE MENETELMILLE

Tässä liitteessä esitellään Eulerin menetelmän, Ralstonin menetelmän ja Klassisen Runge–Kutta-menetelmän käytännön toteutukset MATLAB:lla. Esitettävät funktiot eivät tarkasta käyttäjän antamia argumentteja, sillä tarkoituksena on vain näyttää menetelmien eri vaiheet.

A.1 Eulerin menetelmä

```

1 function [x,y] = euler(f,xspan,y0,h)
2 % EULER ratkaisee alkuarvotehtävän  $y'(x) = f(x,y(x))$  numeerisesti Eulerin
3 % menetelmän avulla.
4
5 % Pisteet, joissa alkuarvotehtävä numeeriset ratkaisut lasketaan.
6 x = xspan(1):h:xspan(2);
7
8 % Alustetaan vektori/matriisi y, johon numeeriset ratkaisut kirjataan. y on
9 % vektori, kun alkuarvotehtävä on skalaarimuotoinen (2.1) ja matriisi, kun
10 % alkuarvotehtävä on vektorimuotoinen (2.7).
11 y = zeros(length(y0),length(x));
12
13 N = length(x) - 1; % askelten määrä
14 y(:,1) = y0;      % alkuehto/-arvo
15
16 % Lasketaan numeeriset ratkaisut jokaisessa pisteessä x.
17 for n = 1:N
18     % numeerinen ratkaisu pisteessä  $x_{n+1}$ , ks. (3.14)
19     y(:,n + 1) = y(:,n) + h * f(x(n),y(:,n));
20 end

```

A.2 Ralstonin menetelmä

```

1 function [x,y] = ralston(f,xspan,y0,h)
2 % RALSTON ratkaisee alkuarvotehtävän  $y'(x) = f(x,y(x))$  numeerisesti
3 % Ralstonin menetelmän avulla.
4
5 % Pisteet, joissa alkuarvotehtävä numeeriset ratkaisut lasketaan.

```

```

6 x = xspan(1):h:xspan(2);
7
8 % Alustetaan vektori/matriisi y, johon numeeriset ratkaisut kirjataan, ks.
9 % Liite A.1 rivit 8-10.
10 y = zeros(length(y0),length(x));
11
12 N = length(x) - 1; % askelten määrä
13 y(:,1) = y0;      % alkuehto/-arvo
14
15 % Lasketaan numeeriset ratkaisut jokaisessa pisteessä x.
16 for n = 1:N
17     % ks. (4.30)
18     k1 = f(x(n),y(:,n));
19     k2 = f(x(n) + 2/3*h,y(:,n) + 2/3*h*k1);
20
21     % numeerinen ratkaisu pisteessä  $x_{n+1}$ 
22     y(:,n + 1) = y(:,n) + h/4 * (k1 + 3*k2); % ks. (4.30)
23 end

```

A.3 Klassinen Runge–Kutta-menetelmä

```

1 function [x,y] = rk4(f,xspan,y0,h)
2 % RK4 ratkaisee alkuarvottehtävän  $y'(x) = f(x,y(x))$  numeerisesti klassisen
3 % Runge-Kutta-menetelmän avulla.
4
5 % Pisteet, joissa alkuarvottehtävä numeeriset ratkaisut lasketaan.
6 x = xspan(1):h:xspan(2);
7
8 % Alustetaan vektori/matriisi y, johon numeeriset ratkaisut kirjataan, ks.
9 % Liite A.1 rivit 8-10.
10 y = zeros(length(y0),length(x));
11
12 N = length(x) - 1; % askelten määrä
13 y(:,1) = y0;      % alkuehto/-arvo
14
15 % Lasketaan numeeriset ratkaisut jokaisessa pisteessä x.
16 for n = 1:N
17     % ks. (4.31)
18     k1 = f(x(n),y(:,n));
19     k2 = f(x(n) + h/2,y(:,n) + h/2*k1);
20     k3 = f(x(n) + h/2,y(:,n) + h/2*k2);
21     k4 = f(x(n) + h,y(:,n) + h*k3);
22
23     % numeerinen ratkaisu pisteessä  $x_{n+1}$ 
24     y(:,n + 1) = y(:,n) + h/6 * (k1 + 2 * k2 + 2 * k3 + k4); % ks. (4.31)
25 end

```

LIITE B: ALKUARVOTEHTÄVIEN RATKAISEMISESSA KÄYTETYT MATLAB-KOODIT

Tässä liitteessä esitellään Esimerkissä 3.3 ja Luvussa 5 käytetyt MATLAB-koodit. Jokaisessa koodissa käytetään Liitteen A funktioita.

B.1 Numeerinen ratkaiseminen Eulerin menetelmällä

```

1 % Ratkaistaan Esimerkin 3.3 alkuarvotehtävä  $y'=y$ ,  $y(0)=1$  numeerisesti eri
2 % askelpituuksia käyttäen. Käytetään ratkaisemiseen Eulerin menetelmää ks.
3 % (3.14).
4
5 % ratkaistava alkuarvotehtävä
6 f = @(~,y) y; %  $y' = y$ , eli  $f(y)=y$ 
7 y0 = 1; % alkuehto/-arvo
8 xspan = [0,1]; % tarkasteltava väli
9
10 % Tarkka ratkaisu  $y = \exp(x)$ .
11 y_tarkka = @(x) exp(x);
12
13 % Numeerinen ratkaisu Eulerin menetelmällä eri askelpituuksilla
14 [x1,y1] = euler(f,xspan,y0,0.5); % h = 0,5
15 [x2,y2] = euler(f,xspan,y0,0.25); % h = 0,25
16 [x3,y3] = euler(f,xspan,y0,0.125); % h = 0,125
17 [x4,y4] = euler(f,xspan,y0,0.01); % h = 0,01
18
19 % Piirretään kuvaaja
20 fplot(y_tarkka,xspan,'k--','LineWidth',1.5) % tarkka ratkaisu
21 hold on
22 % Numeeriset ratkaisut
23 plot(x1,y1,'rx-')
24 plot(x2,y2,'gx-')
25 plot(x3,y3,'bx-')
26 plot(x4,y4,'m-')
27 hold off
28
29 % kuvaajan muotoilua
30 legend('$y(x)=e^x$', '$h=0{,}5$', '$e_N \approx 0{,}4683$', ...
31 '$h=0{,}25$', '$e_N \approx 0{,}2769$', ...
32 '$h=0{,}125$', '$e_N \approx 0{,}1525$', ...

```

```

33     '$h=0{,}01$, $e_N \approx 0{,}0135$', ...
34     'interpreter','latex', 'Location','northwest')
35 xlabel('$x$', 'Interpreter', 'latex')
36 ylabel('$y$', 'Interpreter', 'latex')
37 set(gca, 'TickLabelInterpreter', 'latex')
38 grid on

```

B.2 Alkuarvotehtävän ratkaiseminen: putoava pallo

```

1  % Ratkaistaan alkuarvotehtävä (5.6) Eulerin menetelmällä, Ralstonin
2  % menetelmällä ja Klassisella Runge-Kutta-menetelmällä. Piirretään
3  % ratkaisusta kuvaaja. Lasketaan kokonaisvirheiden itseisarvot jokaisessa
4  % ratkaisupisteessä ja piirretään niistä kuvaaja.
5
6  % Alkuarvotehtävässä olevat parametrit.
7  m = 5; % pallon massa (kg)
8  k = 1; % verrannollisuuskerroin (kg/m)
9
10 % Ratkaistava tehtävä muotoa  $v'=f(v)$ , ks. (5.6)
11 f = @(~,v) 9.81 - k/m*v.^2;
12
13 v0 = 0; % alkuehto  $v(0) = 0$  (m/s)
14 tspan = [0,6]; % tarkasteltava aikaväli
15 h = 0.5; % askelpituus
16
17 % alkuarvotehtävän tarkka ratkaisu, ks (5.7).
18 v = @(t) sqrt(9.81*m/k)*tanh(sqrt(9.81*k/m)*t);
19
20 % numeeriset ratkaisut eri menetelmillä
21 [t,v_euler] = euler(f,tspan,v0,h); % Euler
22 [~,v_ralston] = ralston(f,tspan,v0,h); % Ralston
23 [~,v_rk4] = rk4(f,tspan,v0,h); % Klassinen Runge-Kutta
24
25 % Piirretään kuvaaja: Alkuarvotehtävän numeeriset ratkaisut ja tarkka
26 % ratkaisu samaan kuvaajaan
27 plot(t,v_euler,'rx-')
28 hold on
29 plot(t,v_ralston,'gx-')
30 plot(t,v_rk4,'bx-')
31 fplot(v,tspan,'k')
32 hold off
33
34 % Kuvaajan asetuksia
35 legend('Euler','Ralston','Klassinen Runge--Kutta','tarkka ratkaisu', ...
36     'Interpreter','latex','Location','southeast')
37 xlabel('$t \ (s)$', 'Interpreter', 'latex')
38 ylabel('$v \ (m/s)$', 'Interpreter', 'latex')
39 set(gca, 'TickLabelInterpreter', 'latex')
40 grid on

```

```

41
42 % Kokonaisvirheiden itseisarvot eri menetelmille
43 errors_euler = abs(v(t) - v_euler);
44 errors_ralston = abs(v(t) - v_ralston);
45 errors_rk4 = abs(v(t) - v_rk4);
46
47 % Piirretään kokonaisvirheen itseisarvo t:n funktiona välillä tspan, kun
48 % y-akseli logaritminen.
49 figure
50 semilogy(t,errors_euler,'rx-');
51 hold on
52 semilogy(t,errors_ralston,'gx-');
53 semilogy(t,errors_rk4,'bx-')
54 hold off
55
56 % kuvaajan asetuksia
57 legend('Euler','Ralston','Klassinen Runge--Kutta','Interpreter','latex')
58 xlabel('$t \ (s)$','Interpreter','latex')
59 ylabel('$| e_n | = |v(t_n) - v_n|$', 'Interpreter','latex')
60 set(gca, 'TickLabelInterpreter','latex')
61 grid on

```

B.3 Kokonaisvirheen itseisarvon maksimi eri askelpituuksilla

```

1 % Ratkaistaan alkuarvot tehtävä (5.6) kahdeksalla eri askelpituudella, ks.
2 % Taulukko 5.1. Käytetään Eulerin menetelmää, Ralstonin menetelmää ja
3 % Klassista Runge-Kutta-menetelmää. Lasketaan kokonaisvirheiden
4 % itseisarvojen maksimit eri askelpituuksilla ja piirretään kuvaaja
5 % kokonaisvirheiden itseisarvojen maksimista askelpituuden funktiona.
6
7 % Alkuarvot tehtävässä olevat parametrit.
8 m = 5; % pallon massa (kg)
9 k = 1; % verrannollisuuskerroin (kg/m)
10
11 % Ratkaistava tehtävä muotoa: v'(t)=f(v(t)).
12 f = @(v) 9.81 - k/m*v.^2; % ks. (5.6)
13
14 v0 = 0; % alkuehto v(0) = 0 (m/s).
15 tspan = [0,6]; % tarkasteltava aikaväli 0-6 sekuntia
16
17 % askelpituudet, ks. Taulukko 5.1
18 h = [1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 0.25, 0.5, 1];
19
20 % alkuarvot tehtävän tarkka ratkaisu, ks. (5.7)
21 v = @(t) sqrt(9.81*m/k)*tanh(sqrt(9.81*k/m)*t);
22
23 % Alustetaan vektorit, joihin kirjataan kokonaisvirheiden itseisarvojen
24 % maksimit eri askelpituuksilla
25 max_error_rk4 = zeros(size(h));

```

```

26 max_error_ralston = zeros(size(h));
27 max_error_euler = zeros(size(h));
28
29 % Lasketaan kokonaisvirheen itseisarvon maksimi eri askelpituuksilla
30 for i = 1:length(h)
31     % numeeriset ratkaisut askelpituudella h(i)
32     [t,v_rk4] = rk4(f,tspan,v0,h(i));           % Klassinen Runge-Kutta
33     [~,v_ralston] = ralston(f,tspan,v0,h(i));   % Ralston
34     [~,v_euler] = euler(f,tspan,v0,h(i));      % Euler
35
36     % kokonaisvirheiden itseisarvojen maksimit askelpituudella h(i)
37     max_error_rk4(i) = max(abs(v(t) - v_rk4));
38     max_error_ralston(i) = max(abs(v(t) - v_ralston));
39     max_error_euler(i) = max(abs(v(t) - v_euler));
40 end
41
42 % Piirretään kuvaaja: kokonaisvirheen itseisarvon maksimi askelpituuden
43 % funktiona, molemmat akselit logaritmisia
44 loglog(h,max_error_euler,'rx-')
45 hold on
46 loglog(h,max_error_ralston,'gx-')
47 loglog(h,max_error_rk4,'bx-')
48 hold off
49
50 % kuvaajan asetuksia
51 legend('Euler','Ralston','Klassinen Runge--Kutta','Interpreter','latex',...
52     'Location','southeast')
53 xlabel('$h$', 'Interpreter', 'latex')
54 ylabel('$\max |e_n|, \ n=0,1, \ldots, N$', 'Interpreter', 'latex')
55 set(gca, 'TickLabelInterpreter', 'latex')
56 grid on

```

B.4 Vektorimuotoisen alkuarvotehtävän ratkaisu

```

1 % Ratkaistaan vektorimuotoinen alkuarvotehtävä (5.10)-(5.11) numeerisesti.
2 % Käytetään ratkaisemiseen Klassista Runge-Kutta-menetelmää
3
4 % Alkuarvotehtävässä olevat parametrit.
5 m = 5; % pallon massa (kg)
6 k = 1; % verrannollisuuskerroin kg/m()
7
8 % Ratkaistava tehtävä muotoa  $y' = f(y)$ , ks. (5.10)-(5.11).
9 f = @(~,y) [y(2); 9.81-k/m*y(2).^2];
10
11 y0 = [0;0]; % alkuehto  $y(0)=[0 \ 0]^T$ 
12 tspan = [0,4]; % tarkasteltava aikaväli
13 h = 0.1; % askelpituus
14
15 % numeerinen ratkaisu Klassisella Runge-Kutta-menetelmällä

```

```
16 [t,y] = rk4(f,tspan,y0,h);
17
18 % Piirretään kuvaaja: paikka ja vauhti ajan funktiona
19 plot(t,y(1,:), 'bx-',t,y(2,:), 'rx-')
20
21 % kuvaajan muotoilua
22 legend('paikka $x$', 'vauhti $v$', 'Interpreter', 'latex', 'Location', ...
23        'northwest')
24 xlabel('$t \ (s)$', 'Interpreter', 'latex')
25 ylabel('$x \ (m), \ v \ (m/s)$', 'Interpreter', 'latex')
26 set(gca, 'TickLabelInterpreter', 'latex')
27 grid on
```