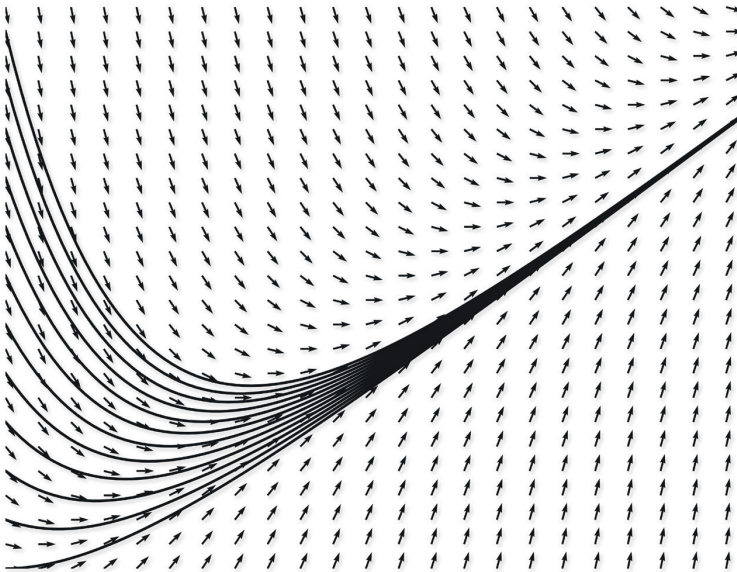


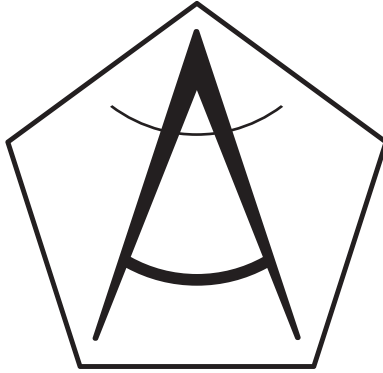
PITKÄN
MATEMATIIKAN
LISÄSIVUT

6

DIFFERENTIAALI-
LASKENTA JA
NUMEERISET
MENETELMÄT



ALEKSIS KOSKI
VILLE TILVIS



PITKÄN MATEMATIIKAN LISÄSIVUT 6

DIFFERENTIAALILASKENTA JA NUMEERISET MENETELMÄT

Aleksis Koski ja Ville Tilvis

Pitkän matematiikan lisäsivut 6: Differentiaalilaskenta ja numeeriset menetelmät

Viipurin Reaalikoulu Oy

Maunulan yhteiskoulu ja Helsingin matematiikkalukio

Matematiikan opetuksen valtakunnallinen kehittämissihteri Harppi

Helsinki 2025

1. Painos

Jotta kirjan käyttö olisi mahdollisimman mutkatonta, sen sisältö on lisensoitu avoimella Creative Commons Nimeä 4.0 Kansainvälinen -käyttöluvalla.



<https://creativecommons.org/licenses/by/4.0/deed.fi>

Kirjaa ja sen osia saa vapaasti levittää ja muokata, kunhan alkuperäiset tekijät mainitaan.

Kannen suunnittelu Salli Kulmala

ISBN 978-952-65761-2-1 (pehmeäkantinen)

ISBN 978-952-65761-3-8 (PDF)

Sisällys

Esipuhe	6
1 Hieman numeriiikasta	7
1.1 Virhe	8
1.2 Virhe laskutoimituksissa	9
1.3 Binääriluvut	11
1.4 Liukuluvut ja toisen asteen yhtälön ratkaisukaava	15
1.5 Numeerinen derivaatta	16
Harjoitustehtäviä	19
2 Taylorin polynomit	22
2.1 Taylorin polynomin idea	22
2.2 Taylorin polynomin määritelmä	25
2.3 Taylorin polynomin virhekaavat	28
Harjoitustehtäviä	32
3 Differentiaalilaskennan perusteita	35
3.1 Raja-arvo	35
3.2 Jatkuvuus	38
3.3 Derivoituvuus	42
Harjoitustehtäviä	45
4 Konveksit funktiot	46
4.1 Funktion konveksisuus	48
4.2 Konveksisuus ja derivaatat	49
4.3 Jensenin epäyhtälö	51

4.4	Optimointiongelmät ja konveksisuus	53
4.5	Usean muuttujan analyysiä	55
	Harjoitustehtäviä	60
5	Numeerinen yhtälönratkaisu	62
5.1	Puolitusmenetelmä	62
5.2	Newtonin menetelmä	64
5.3	Kiintopisteiteraatio	70
5.4	Newtonin dynamiikkaa kompleksitasossa	74
	Harjoitustehtäviä	78
6	Lagrangen interpolaatio	81
6.1	Lagrangen polynomin virhe	84
	Harjoitustehtäviä	88
7	Numeerinen integrointi	89
7.1	Suorakaidesääntö	90
7.2	Puolisuunnikassääntö	92
7.3	Simpsonin sääntö	94
7.4	Numeerisen integroinnin virhearviot	96
7.5	Gaussin integrointisääntö	99
	Harjoitustehtäviä	103
8	Differentiaaliyhtälöt	106
8.1	Separoituvat differentiaaliyhtälöt	109
8.2	Alkuarvo-ongelmat	112
8.3	Suuntakenttä	116
8.4	Lineaariset differentiaaliyhtälöt	117
8.5	Ensimmäisen asteen lineaarinen yhtälö	117
8.6	Toisen asteen vakiokertoiminen lineaarinen yhtälö	119
	Harjoitustehtäviä	124

9 Differentiaaliyhtälön numeerinen ratkaiseminen	127
9.1 Eulerin menetelmä	127
9.2 Runge–Kutta-menetelmä	130
9.3 Differentiaaliyhtälöryhmät	133
9.4 Korkeamman asteen yhtälöt numeerisesti	137
Harjoitustehtäviä	139
Liitteet	141
Liite A: Osittaisintegrointi	141
Liite B: Muuttujanvaihto integraalissa	142
Liite C: Integraalilaskennan yleistetty väliarvolause	144
Liite D: Taylorin polynomin virhekaavojen todistukset	145
Liite E: Newtonin menetelmän suppeneminen	148
Liite F: Eulerin menetelmän virhearvio	150
Ensimmäiset vihjeet	152
Toiset vihjeet	166
Lähteet	182

Esipuhe

Pitkän matematiikan lisäsivut ovat itseopiskelumateriaalia niille, jotka kaipaavat haasteita ja haluavat laajentaa matematiikan osaamistaan lukio-oppimäärän ulkopuolelle. Lisäsivuihin on valittu aihepiirejä, joita lukiossa ei yleensä kohtaa, mutta joihin matematiikasta innostuneen voisi olla kiintoisaa ja palkitsevaa tutustua.

Materiaali sisältää lyhyet teoriaosiot ja runsaasti harjoitustehtäviä, joihin on kattavat vihjeet ja useimpiin toiset vihjeet. Tehtäviä on syytä yrittää aina ensin itse, mutta vihjeet ja toiset vihjeet on tarkoitettu käytettäväksi. Monet tehtävistä ovat huomattavan vaikeita, eikä kannata huolestua, jos osa niistä jää ratkaisematta.

Tämä lisäsivujen kuudes ja näillä näkymin viimeinen osa laajentaa lukiossa käsiteltyä differentiaalilaskennan oppimäärää lisäämällä työkalupakkiin käsitteitä kuten Taylorin polynomit, konveksit funktiot ja differentiaaliyhtälöt. Kaikkia näitä analyttisiä käsitteitä yhdistävät niihin läheisesti liittyvät numeeriset menetelmät ja niiden virhearviointi. Esittelemme kirjassa numeriiikan keskeisiä algoritmeja, kuten Newtonin menetelmän yhtälöiden ratkaisemiseen, numeerisen integroinnin menetelmiä määrättyjen integraalien laskemiseen, ja differentiaaliyhtälöiden ratkaisujen approksimointiin käytettävät Eulerin ja Runge–Kutta-menetelmät. Kokonaisuutena kirjan sisältö tarjoaa rajapintaa sekä tarkkaan differentiaalilaskentaan että käytännön tieteelliseen laskentaan.

Haluamme kiittää kaikkia kollegoita ja opiskelijoita, jotka ovat antaneet kannustavaa palautetta kirjan luonnoksista ja ehdottaneet parannuksia. Avuksi ovat olleet muun muassa Elias Hawkins, Joona Heiskanen, Joona Nikitin, Teo Pohjankoski, Mikko Ylinen ja erityisesti Anu Mäkijärvi, joka oikoluki käsikirjoituksen. Lisäksi Ville Tilvis haluaa kiittää perhettään tuesta ja huomattavasta pitkämielisyydestä kirjaprojektin suhteen.

Kaikista virheistä, puutteista ja parannusehdotuksista toivotaan sähköpostia osoitteeseen ville.tilvis (at) mayk.fi. Antoisia hetkiä lisäsivujen parissa!

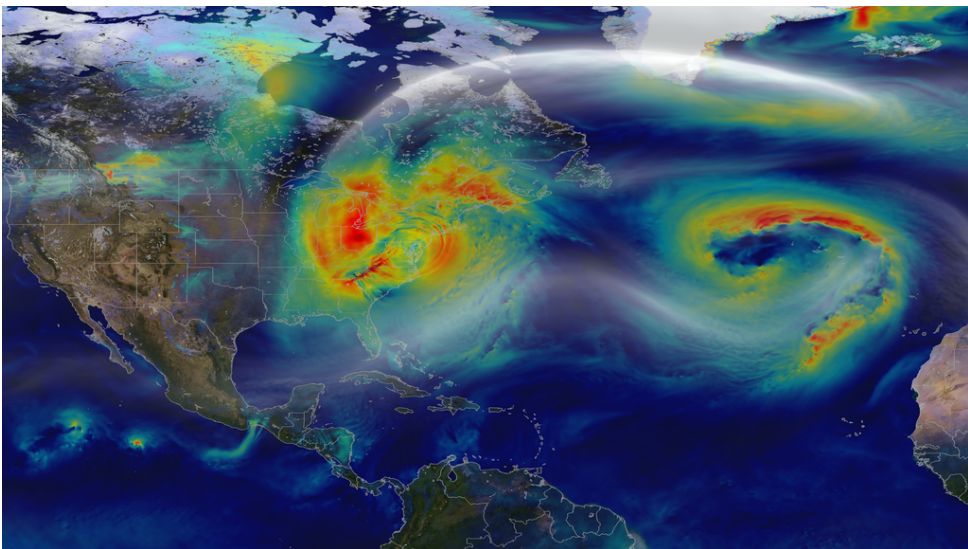
Helsingissä 18.4.2025

Aleksis Koski ja Ville Tilvis

Opettaja! Tehtävien ratkaisut voi tilata osoitteesta ville.tilvis (at) mayk.fi

1 Hieman numeriikasta

Matematiikka on täsmällistä, mutta matematiikan soveltaminen käytäntöön vaatii yleensä likiarvojen ja muuten virheellisten lukujen kanssa työskentelyä. **Numeriikka** on matematiikan ala, joka käsittelee likiarvolaskentaa. Hyvät numeeriset menetelmät tuottavat riittävän tarkkoja likiarvoja riittävän pienellä laskenta-ajalla. Erilaisten algoritmien vakauden ja nopeuden arviointi ja uusien menetelmien kehittäminen ovat numeriikan keskeisiä tehtäviä. Sään ennustaminen, tietokonepelien 3D-maailmojen valaistuksen tehokas laskeminen, tekoälyn kouluttaminen tai hampaan muodon päättely kaksiuotteisista röntgenkuvista ovat esimerkkejä vaativista numeerisista ongelmista.



NASA:n simulaatio hurrikaani Sandyn kehittymisestä, 2012. [25]

Likiarvojen pyörittely voi äkkiseltään kuulostaa puuduttavalta kirjanpidolta, mutta ilahduttavaa kyllä numeriikka tarjoaa runsaasti käyttöä hyvin syvälliselle matemaatiikalle. Mitä enemmän tuntee matematiikkaa, sitä ovelammin ja tehokkaammin laskentaa on mahdollista suorittaa. Esimerkiksi neliöjuurille likiarvon laskeminen on huomattavan tehokasta derivaatan käsitteeseen perustuvalla Newtonin menetelmällä, jota käsitellään Luvussa 5.

1.1 Virhe

Käytössämme on harvoin tarkasti oikeita lukuja. Virheitä voi syntyä monesta lähteestä. Jos tavoitteena on esimerkiksi ennustaa säätä, virhelähteitä ovat

1. **Virheelliset lähtöarvot.** Sääasemien mittaamat lämpötilat ja tuulten nopeudet eivät ole tarkkoja, eivätkä havainnot kattavia.
2. **Mallin virheet.** Fysiikkamalli pyrkii tavoittamaan lopputuloksen kannalta tärkeimmät seikat, eikä huomioi jokaista yksittäistä ilmamolekyyliä.
3. **Numeeristen menetelmien virheet.** Ilmakehän virtauksia kuvaavia differentiaaliyhtälöitä ei osata ratkaista tarkasti, joten niille lasketaan likimääräiset ratkaisut.
4. **Laskennan virhe.** Tietokoneen laskuissa on käytössä rajallinen määrä desimaaleja.

Virhe on väistämätön osa lähes kaikkea laskentaa, joten se täytyy huomioida ja hallita. Määrittelemme virheen seuraavasti:

$$\begin{aligned}\text{Tarkka arvo} &= \text{Likiarvo} + \text{Virhe}, \\ \text{eli} \\ \text{Virhe} &= \text{Tarkka arvo} - \text{Likiarvo}.\end{aligned}$$

Lisäksi kutsumme virheen itseisarvoa nimellä **absoluuttinen virhe** sekä virheen ja oikean arvon osamäärän itseisarvoa nimellä **suhteellinen virhe**.

$$\begin{aligned}\text{Absoluuttinen virhe} &= |\text{Virhe}| \\ \text{Suhteellinen virhe} &= \left| \frac{\text{Virhe}}{\text{Tarkka arvo}} \right|\end{aligned}$$

Jos esimerkiksi **approksimoimme** eli käytämme likiarvoa $\pi \approx 3,1$, virheeksi saadaan $\pi - 3,1 = 0,04159\dots$. Jos taas approksimoimme $\pi \approx 4$, likiarvo on tarkkaa arvoa suurempi, jolloin virhe on negatiivinen, nimittäin $\pi - 4 = -0,8584\dots$. Näiden likiarvojen suhteelliset virheet ovat $\left| \frac{\pi - 3,1}{\pi} \right| \approx 1,32\%$ ja $\left| \frac{\pi - 4}{\pi} \right| \approx 27,32\%$.

Virheen käsitteeseen liittyvä sanasto ei ole täysin vakiintunut. Joissakin yhteyksissä virheen merkki on päinvastainen kuin määrittelimme, ja toisinaan absoluuttista virhettä käytetään virheen synonyyminä. Tässä kirjassa käytämme edellä kiinnitettyjä määritelmiä.

1.2 Virhe laskutoimituksissa

Peruskoulun käyneet ovat todennäköisesti törmänneet sääntöön ”vastaus pyöristetään epätarkimman lähtöarvon mukaan”. Yhteenlaskussa epätarkin arvo määritellään desimaalien perusteella ja kertolaskussa merkitsevien numeroiden perusteella. On hyvä pysähtyä hetkeksi miettimään, miksi näin kannattaa sopia.

Jos esimerkiksi pöydän korkeudeksi on mitattu millien tarkkuudella 73,1 cm, ja sen päällä lepää kirja, jonka paksuudeksi on mitattu mikrometriruuville 2,336 cm, ei ole perustelua väittää kirjan yläreunan olevan 75,436 cm maan yläpuolella, sillä yhteenlaskussa

$$\begin{array}{r} 73,1?? \\ + 2,336 \\ \hline 75,4?? \end{array}$$

kaikki desimaalit ensimmäisen jälkeen ovat epävarmoja. Itse asiassa emme edes tiedä, olisiko oikea likiarvo 75,4 cm vai 75,5 cm.

Yleisemmin voisi merkitä, että jos luvusta a käytetään likiarvoa a' , jonka virhe on Δa , ja vastaavasti $b = b' + \Delta b$, summan $a + b$ likiarvon $a' + b'$ virhe on

$$(a + b) - (a' + b') = (a' + \Delta a + b' + \Delta b) - (a' + b') = \Delta a + \Delta b.$$

Jos virheet Δa ja Δb ovat kovin eri kokoluokkaa, suurempi virhe dominoi, kuten edellisessä kirja ja pöytä -esimerkissä. Jos taas virheet ovat samaa kokoluokkaa, on niiden summakin todennäköisesti samaa kokoluokkaa, ja pahimmassakin tapauksessa vain kaksi kertaa alkuperäisen kokoinen.

Kertolasku ja merkitsevät numerot

Kertolaskussakin pyöristetään epätarkimman lähtöarvon mukaan, mutta nyt tarkkuus määritellään merkitsevien numeroiden perusteella. Tämä on hyvä nyrkki-sääntö, sillä kertolaskussa suhteellinen virhe käyttäytyy likimain samoin kuin tavallinen virhe yhteenlaskussa. Tarkastellaan taas lukuja a ja b , joilla on likiarvot a' ja b' sekä niillä virheet Δa ja Δb . Kun tuloa ab arvioidaan likiarvolla $a'b'$, suhteellinen virhe on

$$\begin{aligned} \left| \frac{ab - a'b'}{ab} \right| &= \left| \frac{ab - (a - \Delta a)(b - \Delta b)}{ab} \right| = \left| \frac{ab - ab + b\Delta a + a\Delta b - \Delta a\Delta b}{ab} \right| \\ &= \left| \frac{b\Delta a}{ab} + \frac{a\Delta b}{ab} - \underbrace{\frac{\Delta a\Delta b}{ab}}_{\approx 0} \right| \approx \left| \frac{\Delta a}{a} + \frac{\Delta b}{b} \right|. \end{aligned}$$

Tässä viimeinen approksimaatio perustuu oletukseen, että virhe on pienempää kokoluokkaa kuin luku itse. Tulon suhteellisen virheen suuruus määräytyy siis likimain tulon tekijöiden (etumerkillä varustettujen) suhteellisten virheiden summan perusteella. Tässäkin epätarkemman lähtöarvon merkitys on suurin.

Vähennyslasku on vaarallinen

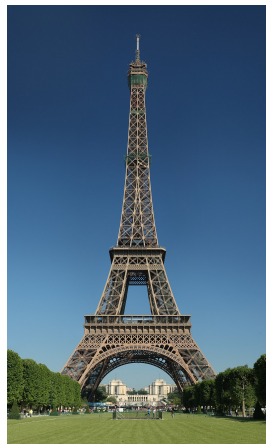
Vähennyslasku on tarkkuuden kannalta vaarallinen laskutoimitus, sillä kun luvut ovat samaa kokoluokkaa, menetetään merkitseviä numeroita ja suhteellinen virhe voi kasvaa rajusti. Kuvitellaan esimerkiksi, että haluaisimme arvioida lämpölaajenemisen vaikutusta Eiffel-tornin korkeuteen mittaamalla yksittäisen rautapalkin pituuden millin tarkkuudella eri lämpötiloissa. Mittaus 0°C lämpötilassa antaa pituudeksi 1,012 m ja mittaus 40°C lämpötilassa 1,013 m. Pituuksien erotukseksi saadaan

$$1,013 \text{ m} - 1,012 \text{ m} = 0,001 \text{ m} = 1 \text{ mm}.$$

Erotuksessa on tarkkuutta jäljellä enää yksi merkitsevä numero, vaikka molemmissa lähtöarvoissa merkitseviä numeroita oli neljä. Todellisuudessa palkkien pituusero olisi alle puoli milliä, joten arvion 1 mm suhteellinen virhe on yli 100 %. Millin tarkkuudella mittaaminen tarkoittaa, että pyöristyksestä riippuen tulos on joko 1 mm tai 0 mm. Saatua tulosta skaalaamalla saataisiin koko 330 metriä korkean tornin pituusvaihteluksi 33 cm, mikä on paljon todellista suurempi arvo. Tilanne ei korjaannu, vaikka tornin uudeksi pituudeksi laskettaisiin verrannon avulla

$$\frac{1,013}{1,012} \cdot 330 \text{ m} = 330,3260 \dots \text{ m},$$

sillä lämpölaajenemisen määrä saadaan kuitenkin erotuksena, jossa katoaa kolme merkitsevää numeroa: $330,3260 \dots \text{ m} - 330 \text{ m} \approx 0,33 \text{ m}$.



Hieman formaalimmin ilmaistuna: Jos tarkoille arvoille a ja b on likiarvot a' ja b' , joiden virheet ovat Δa ja Δb , suhteellinen virhe erotuksessa $a - b$ on

$$\left| \frac{(a - b) - (a' - b')}{a - b} \right| = \left| \frac{a - a' + b - b'}{a - b} \right| = \left| \frac{\Delta a + \Delta b}{a - b} \right|.$$

Tällä lausekkeella ei ole ylärajaa, kun luku a lähestyy lukua b , eli suhteellinen virhe voi olla mielivaltaisen suuri.

1.3 Binääriluvut

Käytämme laskemiseen lukusanoja, joissa kantaluku 10 on erityisessä asemassa. Laskemisen esihistoria on mutkikasta, mutta suurella todennäköisyydellä luvun 10 erityisasema johtuu siitä, että ihmisellä on kymmenen sormea.

Moderni 10-kantainen paikkajärjestelmä lukujen kirjoittamiseen otettiin käyttöön Intiassa 500-luvun lopulla ja se levisi nopeasti ympäri oppinutta maailmaa. Eurooppaan se saapui Espanjan kautta arabialaisen sivistyksen tuomana vuoden 1000 paikkeilla. Keskieurooppalaisten kauppiaiden ja ammatinharjoittajien keskuudessa nämä intialaisarabialaiset luvut syrjäyttivät kömpelöt roomalaiset numerot vasta 1500-luvulla.

Tietokoneissa käytetään pääasiassa **binäärilukuja** eli **parijärjestelmää**, joissa numeromerkkejä on vain kaksi: 1 ja 0. Tämä yksinkertaistaa tiedon fyysistä tallentamista ja laskutoimituksiin tarvittavia algoritmeja. Binääriluvut noudattavat samanlaista paikkajärjestelmän rakennetta kuin kymmenjärjestelmä, kantaluku vain on kaksi. Kymmenjärjestelmän merkintä 5023 tarkoittaa lukua

$$5023 = \begin{array}{c} 10^3 \quad 10^2 \quad 10^1 \quad 10^0 \\ \boxed{5} \quad \boxed{0} \quad \boxed{2} \quad \boxed{3} \end{array} = 5 \cdot 10^3 + 0 \cdot 10^2 + 2 \cdot 10^1 + 3 \cdot 1 \\ = 5000 + 20 + 3.$$

Vastaavasti binääriluku 110111 vastaisi kymmenjärjestelmän lukua

$$110111 = \begin{array}{c} 2^5 \quad 2^4 \quad 2^3 \quad 2^2 \quad 2^1 \quad 2^0 \\ \boxed{1} \quad \boxed{1} \quad \boxed{0} \quad \boxed{1} \quad \boxed{1} \quad \boxed{1} \end{array} = 1 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 1 \\ = 32 + 16 + 4 + 2 + 1 = 55.$$

Sekaannusten välttämiseksi järjestelmän kantaluku voidaan merkitä alaindeksillä: $110111_2 = 55_{10}$.

Desimaalilukujen esittäminen binäärijärjestelmässä noudattaa tuttua rakennetta: pilkun jälkeiset numerot kuvaavat kantaluvun käänteislukujen potenssien määriä. Esimerkiksi luku 13,25 voitaisiin muuttaa binäärijärjestelmään seuraavasti:

$$13,25 = 8 + 4 + 1 + \frac{1}{4} = \begin{array}{c} 2^3 \quad 2^2 \quad 2^1 \quad 2^0 \quad 2^{-1} \quad 2^{-2} \\ \boxed{1} \quad \boxed{1} \quad \boxed{0} \quad \boxed{1} \quad \boxed{0} \quad \boxed{1} \end{array} = 1101,11_2$$

Kokonaisluvut tietokoneissa

Tietokone on äärellinen laite, joten sen muistiin ei voi tallentaa mielivaltaisen suuria lukuja. Perinteinen tapa kokonaislukujen hallinnointiin on varata muistista joka luvulle tietty määrä bittejä. Kukin bitti voi olla ykkönen tai nolla, ja luku

Tyypillinen tapa tallentaa desimaalilukuja tietokoneella on **liukuluku** (englanniksi *floating point number, float*). Liukuluvuissa on kaksi osaa: jokin kantaluvun potenssi osoittamassa kokoluokkaa ja sen kerroin (**mantissa**), jossa on käytössä tietty määrä merkitseviä numeroita.

Tarkastellaan aluksi epärealistisen rajoittunutta systeemiä, jossa mantissassa on käytössä viisi merkitsevää kymmenjärjestelmän numeroa ja eksponentissa yksi numero. Lisäksi tallennetaan luvun ja eksponentin merkki. Systeemin luvut olisivat siis muotoa

$$\boxed{\pm} \boxed{} \boxed{} \boxed{} \boxed{} \boxed{} \cdot 10^{\boxed{\pm} \boxed{}}.$$

Tarkastellaan joitakin esimerkkejä siitä, miten luvut tallentuisivat tähän järjestelmään. Monia lukuja joudutaan pyöristämään, koska merkitseviä numeroita on käytössä vain viisi:

$$\begin{aligned} -5 &= \boxed{-} \boxed{5} \boxed{,} \boxed{0} \boxed{0} \boxed{0} \boxed{0} \cdot 10^{\boxed{+} \boxed{0}} \\ \frac{1}{3} &\approx \boxed{+} \boxed{3} \boxed{,} \boxed{3} \boxed{3} \boxed{3} \boxed{3} \cdot 10^{\boxed{-} \boxed{1}} \\ 654321 &\approx \boxed{+} \boxed{6} \boxed{,} \boxed{5} \boxed{4} \boxed{3} \boxed{2} \cdot 10^{\boxed{+} \boxed{4}} \\ 10^{-15} &\approx \boxed{+} \boxed{0} \boxed{,} \boxed{0} \boxed{0} \boxed{0} \boxed{0} \cdot 10^{\boxed{-} \boxed{9}} = 0 \end{aligned}$$

Tässä järjestelmässä pienin positiivinen luku on $0,0001 \cdot 10^{-9}$ ja toiseksi pienin $0,0002 \cdot 10^{-9}$. Nollan ympäristössä liukulukuja on siis suhteellisen tiheässä, mutta mitä suurempiin lukuihin mennään, eksponentit kasvavat ja liukuluvut ovat yhä harvemmassa. Tämän järjestelmän kaksi suurinta lukua olisivat $9,9998 \cdot 10^9$ ja $9,9999 \cdot 10^9$. Näiden välinen etäisyys on jo 100 000 yksikköä. Merkitsevien numeroiden rajallisuus tuottaa toisinaan yllättäviä ongelmia. Esimerkiksi yhteenlaskun $100\,000 + 1$ kanssa käy hullusti: vastauksen pitäisi olla $100\,001 = 1,00001 \cdot 10^5$, mutta vastaus pyöristyy arvoon $1,0000 \cdot 10^5$. Laskin siis väittää $100\,000 + 1 = 100\,000$.

Oikeasti käytössä olevissa liukuluvuissa eksponentit ovat suurempia ja mantissassa on enemmän merkitseviä numeroita, mutta samat ominaisuudet ja rajoitukset ovat olemassa: merkitseviä numeroita on äärellinen määrä, liukuluvut ovat tiheimmillään nollan lähellä ja harvenevat nolasta etäännyttäessä; järjestelmässä on suurin luku ja pienin positiivinen luku.

1.4 Liukuluvut ja toisen asteen yhtälön ratkaisukaava

Toisen asteen yhtälön $ax^2 + bx + c = 0$ reaaliset ratkaisut, missä $a \neq 0$, saadaan tunnetusti kaavasta

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Liukuluvuilla laskettaessa tämä kaava ei valitettavasti aina toimi. Tarkastellaan vaikkapa yhtälöä

$$3x^2 + 10^{50}x + 5 = 0.$$

Varsin moni laskinohjelmisto laskee sujuvasti

$$x_1 = \frac{-10^{50} - \sqrt{(10^{50})^2 - 4 \cdot 3 \cdot 5}}{2 \cdot 3} \approx -3,333333333333333 \cdot 10^{49}$$

ja

$$x_2 = \frac{-10^{50} + \sqrt{(10^{50})^2 - 4 \cdot 3 \cdot 5}}{2 \cdot 3} \approx 0,0000000000000000.$$

Tässä x_2 ei selvästikään voi olla oikein, eihän $x = 0$ ole yhtälön ratkaisu! Oikea likiarvo olisi $x_2 \approx -5,000 \cdot 10^{-50}$. Miksi laskin siis epäonnistuu?

Ongelma johtuu siitä, että b^2 on paljon suurempi kuin $4ac$. Tällöin pätee

$$\sqrt{b^2 - 4ac} \approx \sqrt{b^2 - 0} = |b|.$$

Kun merkitsevät numerot loppuvat kesken, luvut $\sqrt{b^2 - 4ac}$ ja $|b|$ pyörivät samaan liukulukuun. Ratkaisukaava päättyy muotoon

$$x_1 \approx \frac{-b - |b|}{2a} \quad \text{ja} \quad x_2 \approx \frac{-b + |b|}{2a},$$

jolloin positiivisilla luvuilla b saadaan $x_2 \approx 0$ ja vastaavasti negatiivisella luvulla b saadaan $x_1 \approx 0$, eikä vastauksessa ole yhtään merkitsevää numeroa oikein. Vähennyslasku on vaarallinen!

Korjattu toisen asteen yhtälön ratkaisukaava

Edellä esitetty ongelma yhtälön $ax^2 + bx + c = 0$ ratkaisukaavassa voidaan korjata välttämällä vähennyslaskua ja laskemalla ensin yksi juuri kaavalla

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \quad \text{jos } b \geq 0,$$

tai

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad \text{jos } b < 0.$$

Toinen juuri puolestaan lasketaan kaavalla

$$x_2 = \frac{c}{ax_1},$$

ja näin yhtälön kumpikin ratkaisu saatiin selvitettyä ilman samaa kokoluokkaa olevien lukujen vähennyslaskua.

Toisen asteen yhtälön nolasta poikkeavia ratkaisuja x_1 ja x_2 yhdistävä kaava $x_2 = \frac{c}{ax_1}$ voidaan perustella jakamalla polynomi $ax^2 + bx + c$ tekijöihin sen nolakohtien avulla:

$$ax^2 + bx + c = a(x - x_1)(x - x_2).$$

Kun tekijämuotoisen polynomin sulut kerrotaan auki, saadaan

$$a(x - x_1)(x - x_2) = a(x^2 - x_1x - x_2x + x_1x_2) = ax^2 - \underbrace{a(x_1 + x_2)}_{=b}x + \underbrace{ax_1x_2}_{=c}.$$

Vertaamalla muotoon $ax^2 + bx + c$ nähdään, että

$$b = -a(x_1 + x_2) \quad \text{ja} \quad c = ax_1x_2, \quad \text{eli} \quad x_2 = \frac{c}{ax_1}, \quad \text{kun } x_1 \neq 0.$$

1.5 Numeerinen derivaatta

Toisinaan on tarpeen laskea likiarvo funktion derivaatalle. Voi esimerkiksi olla, että derivaattafunktion lauseke on vaikea muodostaa tai funktiosta tunnetaan vain joitakin arvoja, esimerkiksi jos kyseessä on mittausdataa.

Reaalifunktion f derivaatta f' kohdassa x määritellään erotusosamäärän raja-arvona

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Derivaattaa on siis luontevaa approksimoida erotusosamääränä

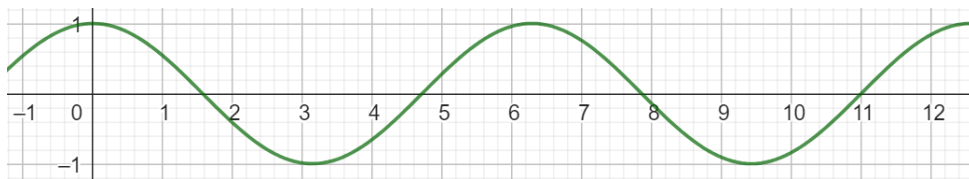
$$f'(x) \approx \frac{f(x+h) - f(x)}{h},$$

missä h on jokin lähellä nollaa oleva luku. Voisi ajatella, että tulos on sitä parempi, mitä pienempi h on. Näin ei kuitenkaan ole, sillä liukulukuvirheen vuoksi laskenta käy hyvin epätarkaksi, kun arvot $f(x+h)$ ja $f(x)$ ovat liian lähellä toisiaan.

Tarkastellaan esimerkiksi funktiota

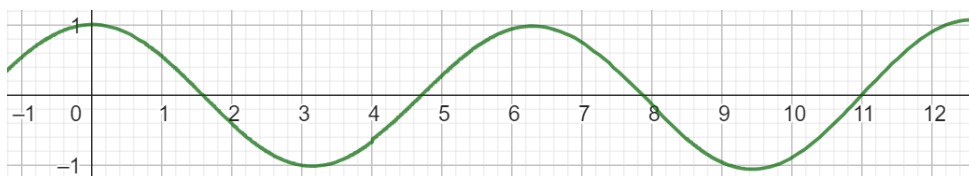
$$C(x) = \frac{\sin(x+h) - \sin x}{h}.$$

Koska sinin derivaattafunktio on kosini, pienillä luvun h arvoilla voisi olettaa pätevän $C(x) \approx \cos x$. Näin käykin, ja kun esimerkiksi $h = 0,01$, GeoGebra 6 -ohjelmalla piirretty funktion C kuvaaja näyttää hyvin paljon kosinifunktion kuvaajalta:



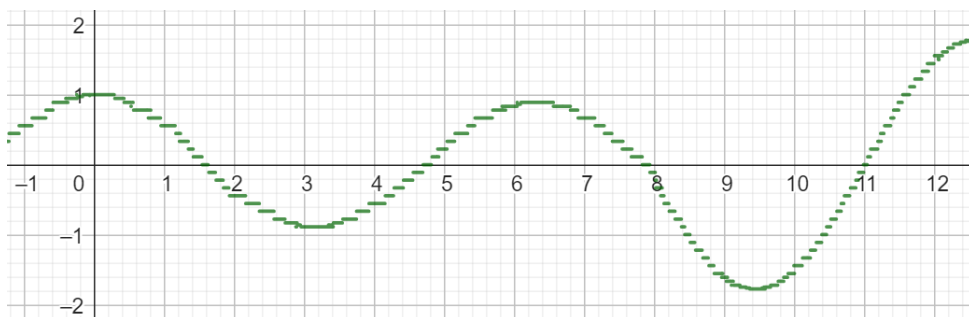
$$h = 10^{-2}$$

Luvun h pientessä yhdennäköisyys kosinifunktion kanssa säilyy arvoon $h = 10^{-14}$ asti hyvänä:



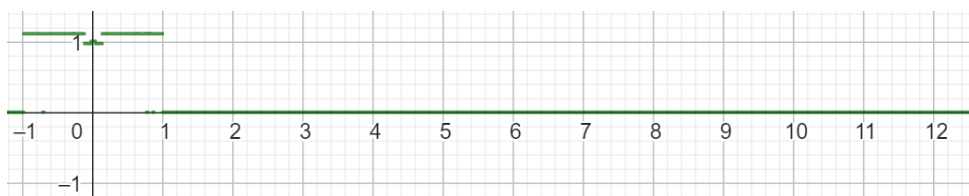
$$h = 10^{-14}$$

Kun $h = 10^{-15}$, liukulukuvirheet alkavat kasautua:



$$h = 10^{-15}$$

Lopulta, kun $h = 10^{-16}$, funktion C lasketut arvot poikkeavat nolasta enää kohdan $x = 0$ läheisyydessä, missä liukulukuja on tiheimmässä.

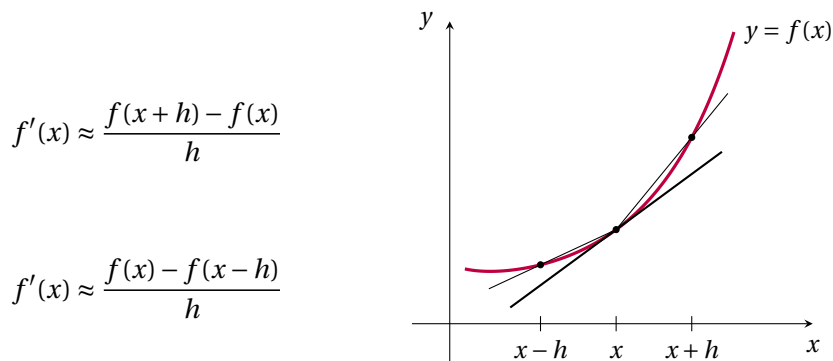


$$h = 10^{-16}$$

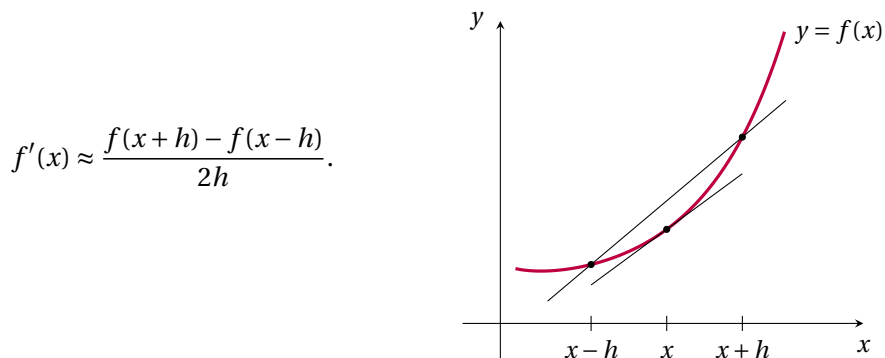
Numeerista derivaattaa laskiessa täytyy siis luonnollisesti kunnioittaa laskentavälineen rajoja.

Keskusdifferenssi

Derivaattaa $f'(x)$ ei yleensä kannata approksimoida erotusosamääränä, jonka toisessa päätepisteessä on kohta x . Esimerkiksi seuraavan kuvan tilanteessa väliä $[x, x+h]$ hyödyntävä arvio antaa liian suuren likiarvon derivaatalle $f'(x)$ ja väliä $[x-h, x]$ käyttävä liian pienen, kun $h > 0$.



Parempi vaihtoehto olisi käyttää **keskusdifferenssiä**, jossa derivaattaa arvioidaan symmetrisellä välillä kohdan x ympäristössä. Keskusdifferenssin kaava on



Teoreettinen perustelu keskusdifferenssin eduista löytyy Taylorin polynomeja käsittelevän Luvun 2 harjoitustehtävistä 20 ja 21.

Harjoitustehtäviä

Tehtävä 1. Kuinka suuri on virhe ja suhteellinen virhe, kun approksimoidaan $\sqrt{5} \approx 2,24$?

Vihje s. 152 →

Tehtävä 2. Matemaattisesti lausekkeen $10^n + 1 - 10^n$ arvo on 1 kaikilla luonnollisilla luvuilla n . Valitettavasti moni laskin on eri mieltä! Esimerkiksi Excel-
taulukkolaskentaohjelma laskee arvon $10^{15} + 1 - 10^{15} = 1$ oikein, mutta väittää, että $10^{16} + 1 - 10^{16} = 0$. Tämä johtuu siitä, että luvun $10^{16} + 1$ esittäminen liukulukuna vaatisi 17 merkitsevää kymmenjärjestelmän numeroa, joten tapahtuu pyöristys $10^{16} + 1 \approx 10^{16}$.

Testaa erilaisten käyttämiesi laskinohjelmistojen tarkkuutta: millä luvun n arvolla lausekkeen $10^n + 1 - 10^n$ arvo poikkeaa ykkösestä? Jos ohjelma käsittelee kokonais- ja liukulukuja eri tavoilla, luku 10 kannattaa syöttää desimaalierottimen kanssa muodossa 10.0. Piirrä ohjelmalla myös funktion $f(x) = 10,0^x + 1 - 10,0^x$ kuvaaja. Miten se käyttäytyy, kun $x > 15$?

Vihje s. 152 →

Tehtävä 3. Tarkastellaan liukulukujen ominaispiirteitä erityisen huonon ja rajoituneen järjestelmän avulla. Mantissassa on käytössä etumerkki ja kaksi kymmenjärjestelmän numeroa sekä eksponentissa etumerkki yksi kymmenjärjestelmän numero. Luvut ovat muotoa

$$\boxed{\pm} \boxed{} \boxed{}, \boxed{} \cdot 10^{\boxed{\pm} \boxed{}}$$

ja järjestelmään tallennetaan siis kaksi etumerkkiä ja kolme numeroa väliltä 0–9. Sovitaan lisäksi, että ensimmäinen numero on nolla vain luvun $0 = 0,0$ tapauksessa, jolloin samaa lukua ei voi tallentaa kahdella eri mantissalla ($1,0 \cdot 10^0$ vs. $0,1 \cdot 10^1$). Jos luku ei mahdu järjestelmään, käytetään normaaleja pyöristyssääntöjä.

- Mikä on suurin luku, joka voidaan tallentaa järjestelmään? Entä toiseksi suurin?
- Mikä on pienin positiivinen luku?
- Miten luku 5 tallennetaan? Entä $1/5$?
- Miten tallentuu $1/3$?
- Mitkä ovat kaksi suurinta peräkkäistä kokonaislukua, siis kokonaisluvut n ja $n + 1$, jotka voidaan tallentaa tarkasti?

- f) Miksi laskut $100 + 2 + 2 + 2 + 2 + 2$ ja $100 + 10$ tuottavat eri lopputuloksen?
- g) Kuinka monta eri lukua järjestelmään voidaan tallentaa?
- h) Kuinka moni näistä luvuista on välillä $]0, 1[$? Entä välillä $]1, 10[$?

Vihje s. 152 →

Tehtävä 4. Binäärilukujen perusteet.

- a) Kirjoita luvut 1–9 binäärilukuina.
- b) Muuta binääriluku 1110 kymmenjärjestelmään.
- c) Esitä kymmenjärjestelmän luku 21 binäärilukuna.
- d) Laske binäärilukujen laskut $1 + 1$, $10 + 1$ ja $100 - 1$.
- e) Binäärilukujen eräs etu on siinä, että perustason laskuoperaatioita on hyvin vähän. Laadi kertotaulu, jossa näkyy kaikkien yksinumeroisten binäärilukujen tulot. Laadi myös vastaava yhteenlaskutaulu.
- f) Laske allekkain binäärilukujen yhteenlasku $100111 + 110110$. Älä muunna kymmenjärjestelmään missään välivaiheessa.
- g) Laske allekkain binäärilukujen kertolasku $1101 \cdot 101$. Älä muunna kymmenjärjestelmään missään välivaiheessa.
- h) Muunna binääriluku 0,011 kymmenjärjestelmän desimaaliluvuksi.

Vihje s. 153 →

Tehtävä 5. Tarkastellaan sivulla 14 esiteltyä binary64-liukulukuformaattia, jota monet ohjelmat käyttävät.

- a) Mikä on pienin positiivinen kokonaisluku, jota järjestelmään ei voi tallentaa tarkasti?
- b) Millaisessa binäärimuodossa luku $-\frac{3}{8}$ tallentuisi?

Vihje s. 153 →

Toinen vihje s. 166 →

Tehtävä 6. Käytössä on likiarvo $\alpha \approx 25,3^\circ$. Miksi luvun $\sin(4000\alpha)$ likiarvosta ei voida sanoa juuri mitään?

Vihje s. 153 →

Toinen vihje s. 166 →

Tehtävä 7. Laske funktion $f(x) = x^x$ derivaatalle $f'(1)$ likiarvo käyttäen

- erotusosamäärää välillä $[1, 1 + h]$
- keskusdifferenssiä välillä $[1 - h, 1 + h]$.

Käytä laskuissa arvoja $h = 0,1$, $h = 0,0001$, $h = 10^{-6}$, $h = 10^{-15}$ ja $h = 10^{-16}$. Selvitä myös derivaatan $f'(1)$ tarkka arvo. Kuinka suuri virhe kussakin tapauksessa on?

Vihje s. 153 →

Tehtävä 8. Tämä tehtävä on versio pitkän matematiikan ylioppilaskokeen tehtävästä syksyltä 2005. Tarkastellaan lauseketta

$$L(x) = \frac{\tan x - \sqrt{3}}{x - \frac{\pi}{3}}.$$

- Laske lauseketta muokkaamatta sille laskimella likiarvo, kun $x = \frac{\pi}{3} + 10^{-3n}$, $n = 1, 2, 3, 4, 5, 6, 7$.
- Määritä $\lim_{x \rightarrow \pi/3} L(x)$ tulkitsemalla lauseke sopivan funktion erotusosamääräksi. Mitä voidaan sanoa a-kohdassa lasketuista likiarvoista?

Vihje s. 154 →

Toinen vihje s. 166 →

Tehtävä 9. Vuonna 1905 julkaistu suppea suhteellisuusteoria ennustaa, että liikkuvan kappaleen aika kuluu hitaammin kuin paikallaan olevan. Tämä on varmistettu kokeellisesti muun muassa maailman ympäri lennätetyillä tarkoilla atomikelloilla [4]. Vauhdilla v etenevän kappaleen kokema aika $\Delta t'$ poikkeaa paikoillaan olevan havaitsijan kokemasta ajasta Δt kaavan

$$\Delta t' = \sqrt{1 - \frac{v^2}{c^2}} \Delta t$$

mukaisesti, missä c on valon nopeus tyhjiössä, eli likimain $3,0 \cdot 10^8$ m/s. Tarkastellaan tutkijaa, joka kävelee etelänavalla (missä Maapallon pyöriminen ei häiritse laskuja) vauhdilla 2 m/s. Kuinka paljon hänen kellonsa jätättää joka sekunti verrattuna paikoillaan olevaan kelloon? On helppo laskea

$$\Delta t - \Delta t' = \Delta t - \sqrt{1 - \frac{v^2}{c^2}} \Delta t = \left(1 - \sqrt{1 - \frac{v^2}{c^2}}\right) \Delta t = \left(1 - \sqrt{1 - \frac{2^2}{300\,000\,000^2}}\right) \cdot 1 \text{ s}.$$

Pahaksi onneksi monet laskimet antavat tästä likiarvoksi esimerkiksi 0,0 s tai $1,78 \cdot 10^{-14}$ s, vaikka oikea likiarvo olisi noin $2,22 \cdot 10^{-17}$ s.

- Kokeile lausekkeen arvon laskemista eri laskinohjelmistoilla. Millaisia tuloksia saat?
- Miten likiarvon $2,22 \cdot 10^{-17}$ s saisi laskettua käsin?

Vihje s. 154 →

Toinen vihje s. 166 →

2 Taylorin polynomit



Brook Taylor (1685–1731)

2.1 Taylorin polynomin idea

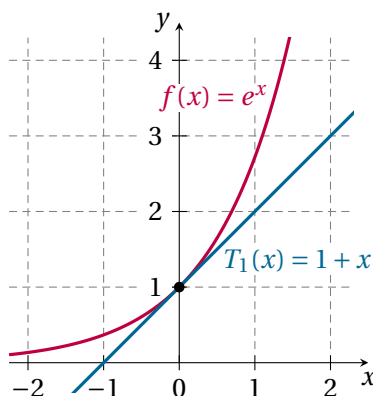
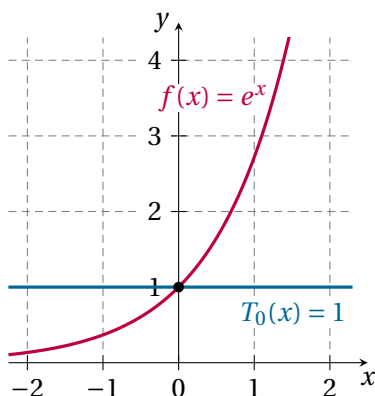
Taylorin polynomi T **approksimoi** jotakin toista funktiota f tietyn pisteen $x = a$ ympäristössä (eng. **approximate**, olla likimain sama). Polynomiapproksimaatio on houkutteleva vaihtoehto monenlaiseen laskentaan, sillä polynomien arvot voidaan laskea peruslaskutoimituksilla ja polynomeilla on yksinkertaisempi teoria kuin monella muulla funktiotyypillä.

Taylorin polynomit T muodostetaan sillä periaatteella, että ne saavat jossakin kohdassa $x = a$

- saman arvon kuin funktio f , eli $T(a) = f(a)$,
- saman derivaatan arvon kuin f , eli $T'(a) = f'(a)$,
- saman derivaatan derivaatan (toisen derivaatan) arvon, eli $T''(a) = f''(a)$, ja niin edelleen.

Tarkastellaan esimerkiksi eksponenttifunktiota $f(x) = e^x$, ja halutaan löytää polynomifunktioita T , jotka saisivat mahdollisimman samansuuruisia arvoja kuin funktio f kohdan $x = 0$ ympäristössä. Koska $f(0) = e^0 = 1$, paras vakiofunktio on $T_0(x) = 1$. Tämä ei luonnollisestikaan ole kovin kummoinen approksimaatio, koska e^x ei suinkaan ole vakio.

Ensimmäisen asteen polynomilla $T_1(x)$ approksimaatio saadaan jo paljon paremmaksi. Valitaan T_1 niin, että sen derivaatta kohdassa $x = 0$ on sama kuin funktiolla f . Koska $f'(x) = e^x$, derivaatta on $f'(0) = e^0 = 1$. Sopiva ensimmäisen asteen polynomi on siis $T_1(x) = 1 + 1 \cdot x$.



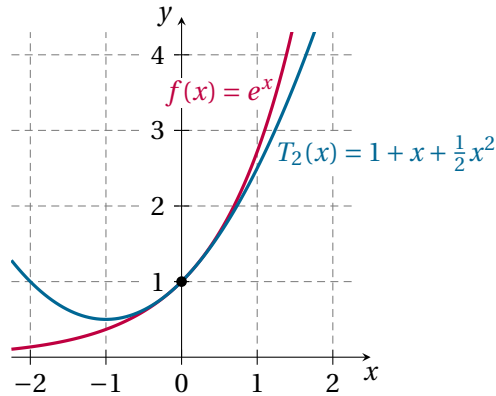
Kun x on lähellä nollaa, $f(x) \approx T_1(x)$. Virheen suuruus kasvaa kuitenkin nopeasti, kun siirrytään kauemmas kohdasta $x = 0$.

x	e^x	$1 + x$	virhe	suhteellinen virhe
0,0	1	1	0	0,00%
0,1	1,01051...	1,1	0,051...	0,47%
0,5	1,64872...	1,5	0,148...	9,02%
1,0	2,71828...	2,0	0,718...	30,10%

Approksimaatiota voidaan parantaa lisäämällä polynomiin $T_1(x)$ vielä yksi termi, jolloin saadaan toisen asteen Taylorin polynomi $T_2(x) = 1 + x + cx^2$. Kerroin c halutaan valita siten, että funktioiden f ja T_2 kuvaajien jyrkkyys muuttuu mahdollisimman samaan tahtiin. Tätä jyrkkyyttä kuvaa derivaatta, ja sen muutosta derivaatan derivaatta eli toinen derivaatta. Valitaan siis kerroin c niin, että $T_2''(0) = f''(0)$. Lasketaan:

$$\begin{array}{ll}
 f(x) = e^x & T_2(x) = 1 + x + cx^2 \\
 f'(x) = D(e^x) = e^x & T_2'(x) = D(1 + x + cx^2) = 1 + 2cx \\
 f''(x) = D(e^x) = e^x & T_2''(x) = D(1 + 2cx) = 2c \\
 f''(0) = e^0 = 1 & T_2''(0) = 2c
 \end{array}$$

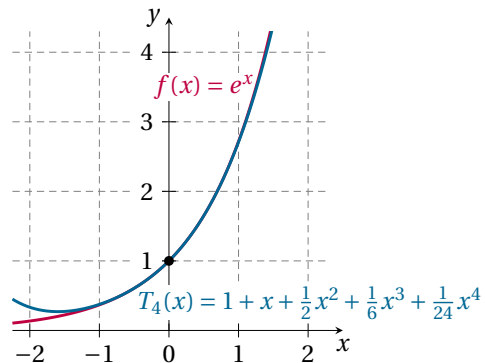
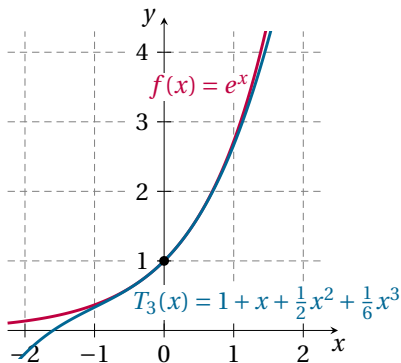
Haluttu ehto $T_2''(0) = f''(0)$ pätee siis, kun $2c = 1$, eli $c = 1/2$. Oikea toisen asteen Taylorin polynomi on siis $T_2(x) = 1 + x + \frac{1}{2}x^2$.



Ero alkuperäisen funktion f ja toisen asteen Taylorin polynomin välillä on jo varsin pieni, kuten myös seuraavasta taulukosta nähdään.

x	e^x	$1 + x + \frac{1}{2}x^2$	virhe	suhteellinen virhe
0,0	1	1	0	0,00%
0,1	1,01051...	1,05	0,00017...	0,016%
0,5	1,64872...	1,625	0,02372...	1,439%
1	2,71828...	2,5	0,21828...	8,030%

Approksimaatiota voidaan entisestään parantaa lisäämällä Taylorin polynomiin korkeamman asteen termejä.



2.2 Taylorin polynomin määritelmä

Määritellään nyt täsmällisesti, mitä tarkoitamme Taylorin polynomeilla.

Taylorin polynomi. Olkoon n positiivinen kokonaisluku. Funktion f kohdassa $x = a$ kehitetty Taylorin polynomi T_n on korkeintaan astetta n oleva polynomifunktio, jolla on kohdassa $x = a$

- sama arvo kuin funktiolla f , eli $T_n(a) = f(a)$,
- sama derivaatan arvo kuin funktiolla f , eli $T'_n(a) = f'(a)$,
- samat korkeamman asteen derivaattojen arvot kertalukuun n asti, eli $T_n^{(k)}(a) = f^{(k)}(a)$, kun $1 \leq k \leq n$.

Määritellään lisäksi erikseen, että nollannen asteen Taylorin polynomi T_0 on vakiofunktio, jolle pätee $T_0(x) = f(a)$.

Taylorin polynomin kaava. Onnekkaisesti Taylorin polynomien muodostamiseen löytyy varsin yksinkertainen kaava.

Lause. Olkoon $n \geq 0$. Funktion f kohdassa $x = a$ kehitetty Taylorin polynomi T_n on muotoa

$$T_n(x) = b_0 + b_1(x-a) + b_2(x-a)^2 + b_3(x-a)^3 + \cdots + b_n(x-a)^n,$$

missä $b_0 = f(a)$ ja kun $1 \leq k \leq n$, kertoimet b_k ovat

$$b_k = \frac{f^{(k)}(a)}{k!}.$$

Taylorin polynomin kaava on siis

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Kaavassa esiintyvä kertoma $n!$ määritellään tulona $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$, missä n on positiivinen kokonaisluku. Esimerkiksi $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. Kaavan voi kirjoittaa tiiviisti

$$T_n(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k,$$

kunhan sovitaan, että $0! = 1$ ja merkitään vielä $f^{(0)}(x) = f(x)$.

Lauseen todistus. Lauseen muotoilun mukaisen polynomin T_n arvo kohdassa $x = a$ on

$$\begin{aligned} T_n(a) &= b_0 + b_1(a-a) + b_2(a-a)^2 + b_3(a-a)^3 + \cdots + b_n(a-a)^n \\ &= b_0 + 0 + 0 + 0 + \cdots + 0, \end{aligned}$$

joten tämä arvo on sama kuin funktiolla f täsmälleen silloin, kun $b_0 = f(a)$. Tutkitaan sitten polynomin T_n derivaattoja. Jokainen derivointi pienentää eksponentteja yhdellä, lisää termeihin kertoimen ja poistaa polynomista yhden termin:

$$\begin{aligned} T_n(x) &= b_0 + b_1(x-a) + b_2(x-a)^2 + b_3(x-a)^3 + \cdots + b_n(x-a)^n \\ T'_n(x) &= 0 + b_1 + 2b_2(x-a) + 3b_3(x-a)^2 + \cdots + nb_n(x-a)^{n-1} \\ T''_n(x) &= 0 + 0 + 2b_2 + 3 \cdot 2b_3(x-a) + \cdots + n(n-1)b_n(x-a)^{n-2} \end{aligned}$$

Lopulta polynomin T_n kertaluvun k derivaatassa $T_n^{(k)}$ ensimmäinen nollasta poikkeava termi on $k(k-1)(k-2) \cdots 3 \cdot 2 \cdot b_k$, eli $k! \cdot b_k$. Kaikissa tätä korkeamman asteen termeissä on tekijänä $(x-a)$, joten kun sijoitetaan $x = a$, saadaan

$$T_n^{(k)}(a) = k! \cdot b_k + 0 + 0 + 0 + \cdots + 0.$$

Jotta tämä olisi sama kuin $f^{(k)}(a)$, täytyy olla $b_k = \frac{f^{(k)}(a)}{k!}$.

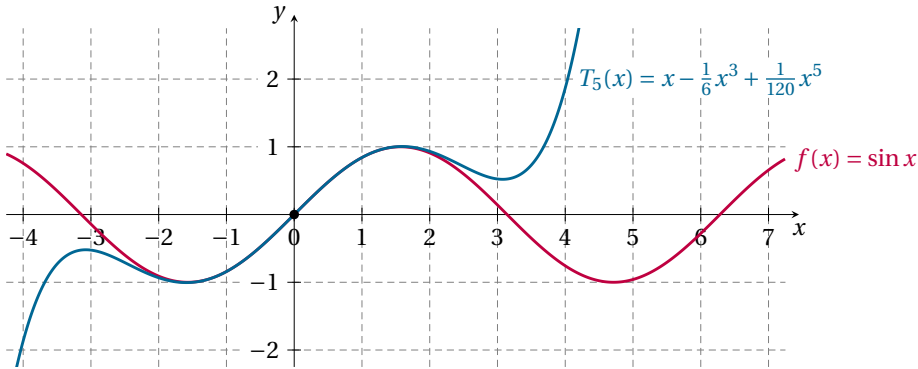
Esimerkki. Lasketaan funktion $f(x) = \sin x$ kohdassa $x = 0$ kehitetty viidennen asteen Taylorin polynomi T_5 . Lasketaan ensin funktion f eri kertalukujen derivaattoja, näiden derivaattojen arvoja kohdassa $x = 0$, ja Taylorin polynomin kertoimia.

$f(x) = \sin x$	$f(0) = \sin 0 = 0$	$b_0 = 0$
$f'(x) = D \sin x = \cos x$	$f'(0) = \cos 0 = 1$	$b_1 = \frac{f'(0)}{1!} = \frac{1}{1!} = 1$
$f''(x) = D \cos x = -\sin x$	$f''(0) = -\sin 0 = 0$	$b_2 = \frac{f''(0)}{2!} = \frac{0}{2!} = 0$
$f'''(x) = D(-\sin x) = -\cos x$	$f'''(0) = -\cos 0 = -1$	$b_3 = \frac{f'''(0)}{3!} = \frac{-1}{3!} = -\frac{1}{6}$
$f^{(4)}(x) = D(-\cos x) = \sin x$	$f^{(4)}(0) = \sin 0 = 0$	$b_4 = \frac{f^{(4)}(0)}{4!} = \frac{0}{4!} = 0$
$f^{(5)}(x) = D \sin x = \cos x$	$f^{(5)}(0) = \cos 0 = 1$	$b_5 = \frac{f^{(5)}(0)}{5!} = \frac{1}{5!} = \frac{1}{120}$

Valitettavasti viidettä derivaattaa ei ole tapana merkitä $f^{(5)}(x)$ vaan tylsästi $f^{(5)}(x)$.

Laskettujen kertoimien avulla voidaan kasata koko polynomi:

$$\begin{aligned} T_5(x) &= b_0 + b_1(x-a) + b_2(x-a)^2 + b_3(x-a)^3 + b_4(x-a)^4 + b_5(x-a)^5 \\ &= 0 + 1 \cdot (x-0) + 0 \cdot (x-0)^2 + \left(-\frac{1}{6}\right) \cdot (x-0)^3 + 0 \cdot (x-0)^4 + \frac{1}{120} (x-0)^5 \\ &= x - \frac{1}{6}x^3 + \frac{1}{120}x^5. \end{aligned}$$

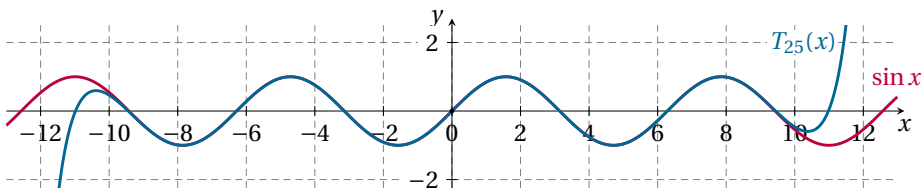


Taylorin polynomien hämmästyttävä tarkkuus

Edellisen esimerkin Taylorin polynomi $T_5(x) = x - \frac{1}{6}x^3 + \frac{1}{120}x^5$ seuraa funktiota $f(x) = \sin x$ yllättävän tarkasti. Mitä tapahtuisi, jos termejä lisättäisiin entisestään? Seuraavat termit on helppo laskea, sillä derivaattalaskuissa havaittiin, että funktiolle $f(x) = \sin x$ pätee $f^{(k+4)}(x) = f(x)$. Korkeampien derivaattojen arvot nollassa, eli luvut $f^{(k)}(0)$ toistuvat siis neljän derivoinnin välein: $1, 0, -1, 0, 1, 0, -1, 0, \dots$. Näin on helppo muodostaa vaikkapa 25. asteen Taylorin polynomi:

$$T_{25}(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \frac{1}{9!}x^9 - \frac{1}{11!}x^{11} + \dots - \frac{1}{23!}x^{23} + \frac{1}{25!}x^{25}.$$

Polynomin T_{25} kuvaaja seuraa sinifunktiota hyvin jo useamman sinifunktion jakson verran. Riittävän suurilla muuttujan x arvoilla korkeimman asteen termi $\frac{1}{25!}x^{25}$ alkaa kuitenkin dominoida, ja polynomin arvot erkaantuvatkin sinin arvoista nopeasti, kun $|x| > 10$.



On lähtökohtaisesti yllättävää, että kohdassa $x = 0$ kehitetty Taylorin polynomi pystyy matkimaan sinifunktiota niin kaukana origosta. Polynomin laatimiseen on kuitenkin käytetty tietoja sinifunktion käyttäytymisestä vain kohdassa $x = 0$ ja sen välittömässä ympäristössä. Mysteeriä voi tarkastella Taylorin polynomin virhekaavan kautta.

2.3 Taylorin polynomin virhekaavat

Kehitetään funktiolle $f(x)$ kohdassa $x = a$ asteen n Taylorin polynomi $T_n(x)$. Määritellään Taylorin polynomin virhetermi $R_n(x)$ luontevasti kaavalla

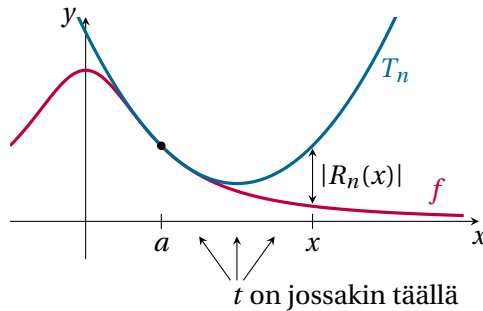
$$R_n(x) = f(x) - T_n(x), \quad \text{eli} \quad f(x) = T_n(x) + R_n(x).$$

Mikäli funktion f asteen n derivaatta $f^{(n)}$ on jatkuva lukujen a ja x rajoittamalla suljetulla välillä ja $f^{(n+1)}$ on olemassa lukujen a ja x rajoittamalla avoimella välillä, pätee

Taylorin polynomin virhekaava

$$R_n(x) = \frac{f^{(n+1)}(t)}{(n+1)!} (x-a)^{n+1},$$

missä t on jokin (lähtökohtaisesti tuntematon) luku lukujen a ja x välissä.



Virhekaavan todistus vaatii hieman enemmän työkaluja kuin tässä on mielekästä käsitellä. Todistuksen pääpiirteet löytyvät liitteestä D sivulta 145 alkaen.

Äkkiseltään virhekaava vaikuttaa hyödyttömältä, koska luku t on tuntematon. Lausekkeen $f^{(n+1)}(t)$ arvolle voidaan kuitenkin usein antaa jonkinlainen ylä- ja alaraja, jolloin virhettä voidaan arvioida.

Esimerkki. Tarkastellaan, mitä virhekaava kertoo funktiolle $f(x) = \sin x$ kohdassa $x = 0$ kehitetystä viidennen asteen Taylorin polynomista T_5 , jonka lauseke on aiemman esimerkin perusteella $T_x = x - \frac{1}{6}x^2 + \frac{1}{120}x^5$. Kuinka hyvä approksimaatio on $\sin(1) \approx T_5(1) = 1 - \frac{1}{6} + \frac{1}{120}$?

Aiemman laskun perusteella $f^{(5)}(x) = \cos x$, joten $f^{(n+1)}(x) = f^{(6)}(x) = \frac{d}{dx} \cos x = -\sin x$. Kohdan $x = 1$ virheeksi saadaan siis

$$R_5(1) = \frac{f^{(n+1)}(t)}{(n+1)!} (x-a)^{n+1} = \frac{-\sin(t)}{(5+1)!} (1-0)^{5+1} = -\frac{\sin(t)}{6!},$$

missä t on jokin luku väliltä $[0, 1]$. Tällä välillä sini saa positiivisia arvoja, joten virhe $R_5(1)$ on negatiivinen. Taylorin polynomin antama likiarvo on siis liian suuri. Koska $-1 \leq \sin t \leq 1$, virheen itseisarvolle saadaan seuraava yläaraja:

$$|f(1) - T_5(1)| = |R_5(1)| = \left| -\frac{\sin t}{6!} \right| = \frac{|\sin t|}{6!} \leq \frac{1}{6!} = \frac{1}{720} = 0,00138\dots$$

Voidaan siis todeta, että approksimaation $\sin(1) \approx 1 - \frac{1}{6} + \frac{1}{120}$ virhe on itseisarvoltaan korkeintaan $1,4 \cdot 10^{-3}$. Huomionarvoista on, ettei tähän arvioon tarvittu tietoa sinifunktion tarkoista arvoista. Todellisuudessa virhe on tätä ylärajaa pienempi, sillä pätee

$$\sin(1) = 1 - \frac{1}{6} + \frac{1}{120} - 0,00019568\dots$$

Itse asiassa virhekaavasta saa tiristettyä virheen itseisarvolle tiukemmankin yläarajan. Sinin Taylorin polynomeissa parittomat termit ovat nollia, joten $x - \frac{1}{6}x^2 + \frac{1}{120}x^5$ on itse asiassa myös sinin kuudennen asteen Taylorin polynomin lauseke. Koska sinifunktion 7. derivaatta on $-\cos x$, kuudennen asteen Taylorin polynomin virheelle saadaan arvio

$$|R_6(1)| = \left| \frac{f^{(6+1)}(t)}{(6+1)!} (1-0)^{6+1} \right| = \left| \frac{-\cos t}{7!} \right| \leq \frac{1}{7!} = 0,00019841\dots,$$

joka on varsin lähellä todellista virhettä.

Taylorin polynomin toinen virhekaava. Vaadimme äskeisessä virhekaavassa erityisesti, että funktiollamme f oli olemassa asteen $n+1$ derivaattafunktio $f^{(n+1)}$. Tämä oletus ei kuitenkaan ole täysin tarpeellinen, sillä ilmeisesti oletusta on mahdollista osoittaa, että Taylorin polynomi approksimoi funktiota hyvin pisteen a läheisyydessä.

Tämä väite voidaan ilmaista toisenlaisena arviona virhetermille $R_n(x)$, jonka muotoilemme nyt. Mikäli funktion f asteen n derivaatta $f^{(n)}$ on jatkuva lukujen a ja x rajoittamalla suljetulla välillä, niin on olemassa sellainen funktio $r_n(x)$, että voimme kirjoittaa

$$R_n(x) = r_n(x)(x-a)^n, \text{ missä } \lim_{x \rightarrow a} r_n(x) = 0.$$

Tämä Taylorin polynomin toinen virhekaava siis ilmaisee yksinkertaisesti sen, että jäännöstermi $R_n(x)$ suppenee kohti nollaa nopeammin kuin asteen n polynomifunktio $(x-a)^n$, sillä virhetermissä esiintyy kertoimena ylimääräinen nollaan suppeneva tekijä $r_n(x)$. Kertoimen suppenemisnopeus tulee riippumaan siitä, kuinka nopeasti asteen n derivaatan $f^{(n)}(x)$ arvot suppenevat kohti arvoa $f^{(n)}(a)$.

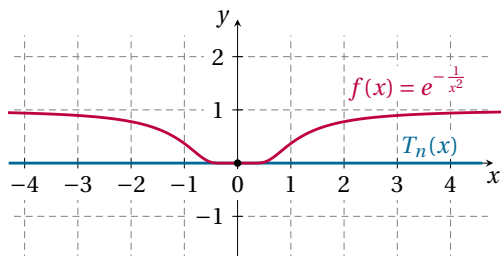
Tätä virhekaavaa on hieman vaikeampi soveltaa käytännön arvioihin verrattuna ensimmäiseen virhekaavaan, mutta se on teoreettisella tasolla vahvempi tulos. Myös toisen virhekaavan todistus löytyy liitteestä D sivulla 146.

Milloin siis Taylorin polynomi toimii kaukana kohdasta $x = a$?

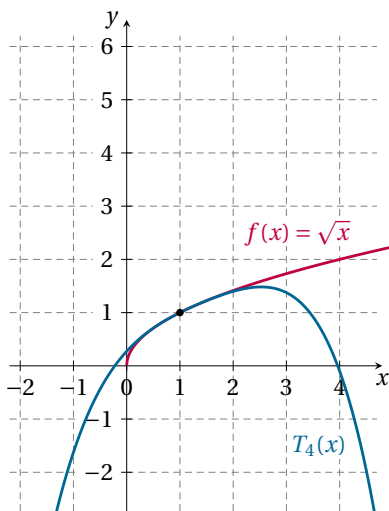
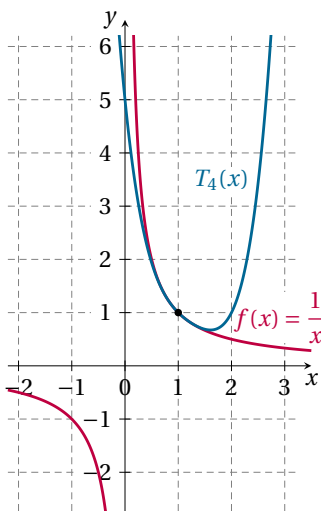
Taylorin polynomit ovat lähtökohtaisesti työkalu arvioimaan funktiota f kehitetävän kohdan $x = a$ välittömässä läheisyydessä, eivätkä välttämättä tarjoa hyviä arvioita kauempana pisteestä a . Koska Taylorin polynomien muodostamiseen hyödynnetään funktion ja sen eri asteiden derivaattojen arvoja ainoastaan kohdassa $x = a$, voi Taylorin polynomi epäonnistua ottamaan huomioon funktion käyttäytymistä kauempana pisteestä a . Esimerkiksi funktiolle

$$f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

löytyy kyllä jokaisen asteen n derivaattafunktio, mutta $f^{(n)}(0) = 0$ kaikilla n . Siis-
pä jokainen tämän funktion Taylorin polynomi on yksinkertaisesti vakiofunktio $T_n(x) = 0$, joten Taylorin polynomit tarjoavat varsin huonon approksimaation funktiolle $f(x)$ kauempana pisteestä $x = 0$.



Vielä suuremman ongelman tuottavat funktiot, jotka eivät ole derivoituvia kohdan $x = a$ läheisyydessä. Esimerkiksi kohdassa $x = 1$ kehitetyt Taylorin polynomit eivät voi matkia rationaalifunktion $f(x) = \frac{1}{x}$ käyttäytymistä kohdan $x = 0$ ympäristössä. Neliöjuurifunktion derivaatta puolestaan saa mielivaltaisen suuria arvoja kohdan $x = 0$ ympäristössä, mitä polynomien on vaikea jäljitellä.



On kuitenkin erikoistapauksia, joissa riittävän korkean asteen Taylorin polynomi tuottaa edelleen varsin hyvän arvion funktiosta vaikka piste x valittaisiin kauempaa. Näin tapahtuu pääasiassa silloin, kun funktion eri asteiden derivaatat eivät kasva liian nopeasti.

Tarkastellaan nyt funktiota f ja sen kohtaan $x = a$ kehitettyjä eri asteiden n Taylorin polynomeja T_n . Jotta approksimaatio $f(x) \approx T_n(x)$ olisi hyvä myös pisteessä x , virhetermin

$$R_n(x) = f^{(n+1)}(t) \cdot \frac{(x-a)^n}{(n+1)!}$$

tulisi olla lähellä nollaa. Kiinnitettyä arvoa x tarkasteltaessa voidaan ajatella lauseke $x - a$ vakiona, joten polynomin astetta n kasvatettaessa lauseke $(n+1)!$ kasvaa paljon nopeammin kuin $(x-a)^n$, kunhan $n+1 > x-a$. Termi $\frac{(x-a)^n}{(n+1)!}$ kutistuu siis kohti nollaa.

Jollekin kiinteälle arvolle x approksimaatiosta $f(x) \approx T_n(x)$ saadaan siis kuinka tarkka hyvänsä käyttämällä riittävän korkean asteen Taylorin polynomia, kunhan termi $f^{(n+1)}(t)$ kasvaa hitaammin kuin termi $\frac{(x-a)^n}{(n+1)!}$ pienenee. Näin tulee tapahtumaan monilla tutuilla funktioilla.

Esimerkiksi sinin tapauksessa $f^{(n+1)}(t)$ pysyy rajoitettuna, sillä $|f^{(n+1)}(t)| \leq 1$ kaikilla $n \geq 1$. Sinin approksimointi halutulla tarkkuudella kuinka suurella välillä $[-x, x]$ tahansa on siis mahdollista, kunhan valitaan lukuun x sopiva polynomin aste n . Tähän liittyy harjoitustehtävä 18.

Lisähuomio: Mainittakoon, että edellä todettujen seikkojen vuoksi sinifunktion voi kuvata tarkasti *äärettömällä* summalla Taylorin polynomien termejä, eli niin sanotulla **Taylorin sarjalla**:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \cdots = \sum_{k=0}^{\infty} \left(\frac{(-1)^k \cdot x^{2k+1}}{(2k+1)!} \right).$$

Emme tässä tekstissä käsittele äärettömiä summia tai Taylorin sarjoja tarkemmin.

Harjoitustehtäviä

Tehtävä 10.

- a) Laske funktiolle $f(x) = e^x$ kohdassa $x = 0$ kehitetty 5. asteen Taylorin polynomi $T_5(x)$.
- b) Laske $T_5(1)$. Miksi se on kelvollinen likiarvo Neperin luvulle e ?

Vihje s. 154 →

Toinen vihje s. 166 →

Tehtävä 11. Laske käsin likiarvo luvulle $\sqrt{5}$ käyttäen funktion $f(x) = \sqrt{x}$ kohdassa $x = 4$ kehitettyä toisen asteen Taylorin polynomia. Anna vastaus murtolukuna.

Vihje s. 154 →

Toinen vihje s. 167 →

Tehtävä 12. Muodosta funktion $f(x) = \cos x$ kohdassa $x = 0$ kehitetty toisen asteen Taylorin polynomi $T_2(x)$. Vertaa funktioiden f ja T_2 kuvaajia.

Vihje s. 154 →

Toinen vihje s. 167 →

Tehtävä 13. Tarkastellaan kahta tapaa approksimoida logaritmifunktiota:

- a) Muodosta funktiolle $f(x) = \ln(x + 1)$ toisen asteen Taylorin polynomi kohdassa $x = 0$.
- b) Muodosta funktiolle $g(x) = \ln x$ toisen asteen Taylorin polynomi kohdassa $x = 1$.

Mitä samaa ja mitä erilaista näissä Taylorin polynomeissa on?

Vihje s. 155 →

Toinen vihje s. 167 →

Tehtävä 14. Tehtävässä 10 funktiota $f(x) = e^x$ approksimoitiin 5. asteen Taylorin polynomilla $T_5(x)$, joka oli kehitetty kohdassa $x = 0$. Laske Taylorin polynomin virhekaavan avulla jokin yläraja virheelle $f(x) - T_5(x)$, kun $x = 1$.

Vihje s. 155 →

Toinen vihje s. 167 →

Tehtävä 15. Miten approksimaatiot

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x \quad \text{ja} \quad \sqrt[3]{1+x} \approx 1 + \frac{1}{3}x$$

liittyvät Taylorin polynomeihin? Antavatko approksimaatio liian suuria vai liian pieniä arvoja? Millaisilla luvun x arvoilla virhe on pieni?

Vihje s. 155 →

Toinen vihje s. 167 →

Tehtävä 16. Muodosta funktion $f(x) = |x|$ kolmannen asteen Taylorin polynomi kohdassa $x = 1$. Miksi Taylorin polynomit eivät toimi hyvin itseisarvofunktion kanssa?

Vihje s. 155 →

Toinen vihje s. 168 →

Tehtävä 17. Kun x on lähellä nollaa oleva, radiaaneissa mitattu kulma, pätee varsin hyvä approksimaatio

$$\sin x \approx x.$$

Osoita Taylorin polynomin virhekaavalla, että approksimaation virheen itseisarvo on korkeintaan $\frac{|x|^3}{6}$.

Vihje s. 155 →

Toinen vihje s. 168 →

Tehtävä 18. Kuvitellaan, että haluat kirjoittaa tietokoneohjelman, joka osaa laskea trigonometrisille funktioille likiarvoja käyttäen vain peruslaskutoimituksia. Eräs (joskaan ei suinkaan ainoa tai paras) tapa olisi käyttää Taylorin polynomeja, saahan polynomien arvot laskettua yhteen-, vähennys- ja kertolaskuilla.

Kehitetään siis funktiolle $f(x) = \cos x$ Taylorin polynomi kohdassa $x = 0$. Kuinka korkean asteen Taylorin polynomia täytyy käyttää, jotta kaikilla muuttujan x arvoilla välillä $[-\pi, \pi]$ virhe olisi varmasti itseisarvoltaan korkeintaan 10^{-8} ?

Vihje s. 156 →

Toinen vihje s. 168 →

Tehtävä 19.

Olkoon n positiivinen kokonaisluku. Muodosta funktioille $f(x) = e^x$, $g(x) = \cos x$ ja $h(x) = \sin x$ kohdassa $x = 0$ kehitetyt asteen n Taylorin polynomit $E(x)$, $C(x)$ ja $S(x)$. Osoita, että pätee

$$E(ix) = C(x) + i \cdot S(x),$$

missä i on kompleksilukujen imaginaariyksikkö, jolle pätee $i^2 = -1$.

Huomautus. Tämä tulos on jokseenkin läheistä sukua kompleksifunktioiden Eulerin kaavalle $e^{ix} = \cos x + i \sin x$.

Vihje s. 156 →

Toinen vihje s. 169 →

Tehtävä 20. Funktion f derivaattaa kohdassa x voidaan approksimoida erotusosamäärällä:

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

missä h on lähellä nollaa oleva positiivinen tai negatiivinen luku. Oletetaan, että toinen derivaatta f'' on jatkuva lukujen x ja $x+h$ muodostamalla suljetulla välillä, ja että tällä välillä toisen derivaatan itseisarvon suurin arvo on $|f''(x)| = M$. Osoita, että approksimaation absoluuttinen virhe on korkeintaan $\frac{1}{2}M|h|$.

Vihje s. 156 →

Toinen vihje s. 169 →

Tehtävä 21. Tehtävän 20 epäsymmetristä kaavaa parempi tapa approksimoida derivaattaa on keskusdifferenssi

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

missä h on pieni positiivinen luku. Oletetaan, että välillä $[x-h, x+h]$ kolmas derivaatta f''' on jatkuva ja sen itseisarvolle pätee $|f'''(t)| \leq M$. Osoita, että tällöin keskusdifferenssin absoluuttinen virhe on korkeintaan $\frac{1}{6}Mh^2$. Tämä on merkittävä parannus epäsymmetrisen erotusosamäärän tapaukseen nähden: kun h on lähellä nollaa, h^2 on paljon pienempi kuin $|h|$.

Vihje s. 156 →

Toinen vihje s. 170 →

3 Differentiaalilaskennan perusteita

Esitämme tässä luvussa matematiikan yleissivistystä ja tulevia lukuja varten pari keskeistä määritelmää ja tulosta, jotka liittyvät jatkuvuuteen ja derivoituvuuteen. Tulemme tässä luvussa myös soveltamaan tarkempaa matemaattista täsmällisyyttä, ja antamaan tarkat määritelmät käsitteille kuten **raja-arvo** ja **jatkuvuus**, sillä näitä ei aina käydä läpi lukion oppimäärässä.

Yleisesti tämän kirjan tavoite ei ole tähdätä täydelliseen matemaattiseen tarkkuuteen, sillä se menisi osittain myös päällekkäin yliopistomatematiikan oppimäärän kanssa, mutta tämä luku on suunnattu lukijalle, joka kaipaa hieman täsmällisempää käsittelyä lukiossa esitetyille analyysin peruskäsitteille. Tässä luvussa käsiteltäviä todistuksia ei tarvitse käydä läpi tulevien lukujen sisällön ymmärtämiseksi.

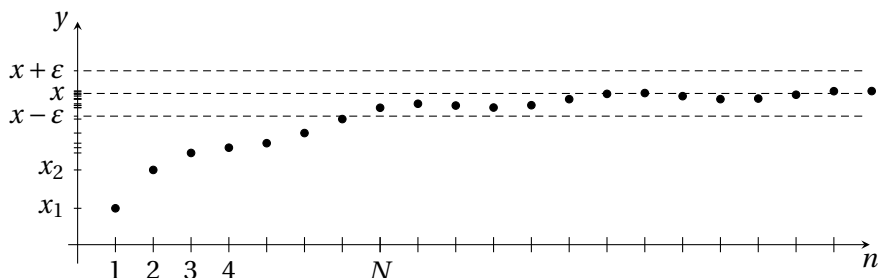
3.1 Raja-arvo

Aloitamme lukujonon raja-arvon täsmällisestä määritelmästä.

Lukujonon raja-arvon määritelmä

Reaalilukujonolla (x_n) on **raja-arvo** x , merkitään $x = \lim_{n \rightarrow \infty} x_n$, mikäli kaikilla $\varepsilon > 0$ on olemassa positiivinen kokonaisluku N siten, että

$$|x_n - x| < \varepsilon \quad \text{kaikilla } n \geq N.$$



Ensimmäistä kertaa nähtynä määritelmä saattaa näyttää melko vieraalta tai tekniseltä, mutta sillä on luonnollinen tulkinta. Määritelmä väittää, että miten tahansa pieni etäisyydsluku $\varepsilon > 0$ valitaankin, niin jostakin, mahdollisesti suuresta,

indeksistä N lähtien lukujonomme päättyy alle luvun ε etäisyydelle raja-arvostaan. Erityishuomio on, että indeksi N riippuu luvun ε valinnasta. Siis mitä pienempi ε valitaan, sitä lähemmäksi jonon jäsenet saadaan raja-arvosta etenemällä jonossa riittävän pitkälle.

Lukujonon raja-arvon määritelmää voi ajatella pelinä, jossa vastustaja antaa pienen etäisyyden $\varepsilon > 0$, ja pelaajan on tarkoitus löytää indeksi N , josta lähtien jono pysyy alle etäisyyden ε päässä väitetyistä raja-arvostaan. Mikäli pelaaja voittaa aina, niin raja-arvo on väitetty luku. Jos taas vastustaja keksii sellaisen luvun ε , jolla jono aina toisinaan karkaa pidemmälle väitetyistä raja-arvosta, niin määritelmä ei toteudukaan ja raja-arvoa ei ole olemassa.

Esimerkki. Olkoon (x_n) jono $x_n = \frac{\sin(n)}{\sqrt{n}}$, missä $n = 1, 2, 3, \dots$. Väitämme, että jonon (x_n) raja-arvo on 0. Todistetaan tämä käyttämällä raja-arvon määritelmää. Määritelmää käyttäessä todistus alkaa aina seuraavalla virkkeellä: Olkoon $\varepsilon > 0$. Kiinnitämme siis mielivaltaisen positiivisen luvun ε , ja pyrimme nyt löytämään sopivan luvun N , jolla määritelmän ehto

$$|x_n - x| < \varepsilon \quad \text{kaikilla } n \geq N$$

toteutuu. Epäyhtälö $|x_n - x| < \varepsilon$ saa nyt muodon $\frac{|\sin(n)|}{\sqrt{n}} < \varepsilon$. Koska sinifunktion arvot on rajattu välille $[-1, 1]$, pätee $|\sin(n)| \leq 1$ kaikilla n . Siispä riittäisi tutkia epäyhtälöä $\frac{1}{\sqrt{n}} < \varepsilon$, joka voidaan myös kirjoittaa muotoon $n > \frac{1}{\varepsilon^2}$. Tämä näyttää jo melko hyvältä, sillä varmasti riittävän isoilla luvuilla n tällainen epäyhtälö pätee. Pääsemme tässä itse asiassa soveltamaan erästä reaalilukujen perusominaisuutta, niin sanottua **Arkhimedeen lausetta**.

Arkhimedeen lause

Kun r on mikä tahansa reaaliluku, niin on olemassa kokonaisluku N siten, että $N > r$.

Erityisesti jos valitaan $r = \frac{1}{\varepsilon^2}$, niin löytyy sellainen (positiivinen) kokonaisluku N , että $N > \frac{1}{\varepsilon^2}$. Tämä N kelpaa meille, sillä nyt jos $n \geq N$, niin

$$|x_n - x| = \frac{|\sin(n)|}{\sqrt{n}} \leq \frac{1}{\sqrt{n}} \leq \frac{1}{\sqrt{N}} < \frac{1}{\sqrt{\varepsilon^{-2}}} = \varepsilon.$$

Siis väite on todistettu, koska löysimme jokaiselle $\varepsilon > 0$ riittävän suuren indeksin N , josta lähtien jonon jäsenet pysyvät alle etäisyyden ε päässä raja-arvostaan.

Funktion raja-arvo

Raja-arvosta on myös funktioille oma versio, jota kutsutaan **funktion raja-arvoksi**. Ideatasolla funktion raja-arvossa on kyse siitä, mitä lukua annetun funktion $f(x)$ arvot lähestyvät, kun pisteen x annetaan lähestyä jotain kiinnitettyä pistettä x_0 .

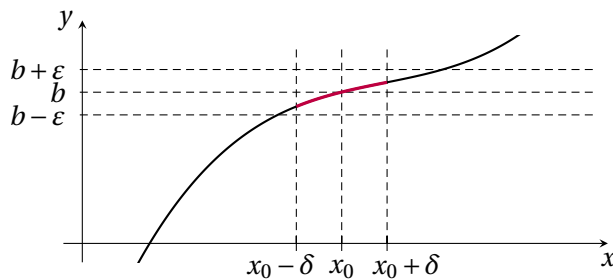
Käsitlemme tässä funktioita $f: I \rightarrow \mathbb{R}$, jotka ovat määriteltyjä jossain reaalilukujen osajoukossa I ja saavat reaalilukuarvoja. Tässä funktion määrittelyjoukko I voi esittää koko reaalilukujen joukkoa $\mathbb{R}$ tai vaikkapa avointa, puoliavointa tai suljettua väliä. Sallimme myös yleisemmät tilanteet, mutta tässä tekstissä keskitymme edellä mainittuihin perustapauksiin.

Funktion raja-arvon määritelmä

Funktiolla $f(x)$ on **raja-arvo** y pisteessä x_0 , merkitään $y = \lim_{x \rightarrow x_0} f(x)$, mikäli kaikilla $\varepsilon > 0$ on olemassa luku $\delta > 0$ siten, että

$$|f(x) - y| < \varepsilon, \quad \text{kun } 0 < |x - x_0| < \delta.$$

Myös tämä määritelmä voi vaikuttaa aluksi hieman hämärältä. Oleellisesti ottaen se sanoo, että mikä tahansa pieni etäisyys $\varepsilon > 0$ valitaankin, niin funktion arvot, eli luvut $f(x)$, saadaan enintään etäisyyden ε päästä sen raja-arvosta y , kunhan piste x valitaan riittävän läheltä pistettä x_0 (eli $|x - x_0| < \delta$ jollain pienellä luvulla $\delta > 0$). Erityisesti δ riippuu luvun ε valinnasta. Määritelmä antaa siis matemaattisesti täsmällisen tavan ilmaista sen, että $f(x)$ on lähellä arvoa y , kun x on lähellä pistettä x_0 .



Tyypillisesti piste x_0 on jokin piste funktion määrittelyjoukossa I , mutta tarkasti katsottuna yllä oleva määritelmä ei vaadi, että funktio f olisi määritelty pisteessä x_0 . Tämä ilmenee ehdosta $0 < |x - x_0| < \delta$, jonka mukaan $x \neq x_0$, ja piste x on korkeintaan etäisyyden δ päässä pisteestä x_0 .

Teemme vielä pienen täsmennyksen määritelmään: Riittää, että määritelmän ehto pätee niille välin $]x_0 - \delta, x_0 + \delta[$ pisteille $x \neq x_0$, jotka kuuluvat funktion määrittelyjoukkoon I . Ei haittaa, vaikka f ei olisi määritelty kaikissa pisteissä pisteen x_0

lähellä, mutta vaadimme, että vaikka δ olisi kuinka pieni, pisteen x_0 ympäristöstä löytyy aina pisteitä x , joissa $f(x)$ on määritelty. Tarkka termi tällaiselle pisteelle x_0 on määrittelyjoukon I **kasaantumispiste**, mutta emme mene tässä syvemmälle yksityiskohtiin.

Riittää muistaa, että jos funktio f on määritelty esimerkiksi avoimella välillä $I =]a, b[$, niin määritelmä antaa tavan puhua funktion f raja-arvosta myös päätepisteissä a ja b , mikäli tällainen raja-arvo on olemassa.

Esimerkki. Funktio $f(x) = \frac{x^2 - 1}{x - 1}$ ei ole määritelty pisteessä $x_0 = 1$, mutta väitämme silti, että sillä on pisteessä 1 olemassa raja-arvo $\lim_{x \rightarrow 1} f(x) = 2$.

Kuten lukujonon raja-arvossa, tämän asian todistaminen alkaa virkkeellä: Olkoon $\varepsilon > 0$. Haluamme nyt löytää riittävän pienen luvun δ , että ehto

$$|f(x) - 2| \leq \varepsilon, \quad \text{kun } 0 < |x - 1| \leq \delta$$

toteutuu. Sieventämällä saadaan, kun $x \neq 1$,

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

Siispä $|f(x) - 2| = |x + 1 - 2| = |x - 1|$. Nyt halutaan valita luku $\delta > 0$ niin, että ehdosta $|x - 1| < \delta$ seuraa ehto $|f(x) - 2| = |x - 1| < \varepsilon$. Mutta tähän selvästi riittää valinta $\delta = \varepsilon$, sillä silloin nämä ovat sama epäyhtälö. Siispä väite on todistettu. Tiivistäen, näytimme, että kun argumentin x arvot ovat etäisyyden δ päässä pisteestä $x_0 = 1$, niin myös funktion arvot ovat annetun etäisyyden ε päästä funktion väitetystä raja-arvosta 2.

3.2 Jatkuvuus

Funktion raja-arvon käsitteen avulla voimme määritellä myös luonnollisella tavalla funktion **jatkuvuuden**. Jos x_0 on piste funktion määrittelyjoukossa, niin sanomme, että funktio f on **jatkuva pisteessä** x_0 mikäli $f(x_0) = \lim_{x \rightarrow x_0} f(x)$. Jatkuvuus tarkoittaa siis yksinkertaisesti sitä, että funktio saa pisteessä x_0 raja-arvokseen oman arvonsa $f(x_0)$. Funktio on **jatkuva**, mikäli se on jatkuva jokaisessa määrittelyjoukon pisteessään.

Jatkuvuudesta puhuttaessa esiintyy usein myös ilmaus ”jatkuva suljetulla välillä $[a, b]$ ”, joka saattaa olla hieman harhaanjohtava. Käytännön syistä tämä virke tulee tulkita väitteenä, että tutkittava funktio f on jatkuva, kun sen määrittelyjoukoksi asetetaan suljettu väli $[a, b]$. Tämä sopimus ratkaisee epäselvyyden tapauksissa, joissa funktio f on alunperin määritelty myös suljetun välin $[a, b]$ ulkopuolella ja käyttäytyy epäjatkovasti, kun välin $[a, b]$ päätepisteitä lähestytään välin ulkopuolelta.

Jatkuvilla funktioilla on paljon näppäriä ominaisuuksia, joihin haluamme pureutua seuraavaksi, mutta käsitellään ensin vielä yksi tärkeä reaalilukuja koskeva peruskäsite.

Supremum ja infimum

Olkoon A joku reaaliluvuista koostuva joukko. Joukolla A on **yläraja** M' , jos luvulle M' pätee

$$x \leq M' \text{ kaikilla luvuilla } x \text{ joukossa } A. \quad (1)$$

Jos joukolla on yläraja, niitä on aina enemmän kuin yksi, sillä myös kaikki jotakin ylärajaa M' suuremmat luvut ovat ylärajoja. Joukon A **supremum** M on kaikista ylärajoista pienin, eli $M \leq M'$ kaikilla luvuilla M' , jotka toteuttavat ehdon (1).

Supremumia merkitään $M = \sup A$. Vastaavalla tavalla voidaan määritellä myös suurin alaraja eli **infimum**, joka on suurin luku m , joka toteuttaa ehdon $m \leq x$ kaikilla $x \in A$. Joukon A infimumia merkitään $\inf A$.

Supremum ja infimum ovat jossakin mielessä yleistyksyet käsitteille maksimi ja minimi. Jos joukolla on maksimi, niin se on aina myös joukon supremum. Mutta esimerkiksi joukolla $A = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots\}$ ei ole olemassa maksimia, sillä jokaisella sen alkiolla $\frac{n}{n+1}$ löytyy joukosta aina suurempi alkio $\frac{n+1}{n+2}$. Voidaan kuitenkin osoittaa, että joukolla A on supremum $\sup A = 1$.

Keskeinen reaalilukujen supremumiin liittyvä ominaisuus on seuraava perustulos.

Reaalilukujen täydellisyysaksiooma

Jokaisella epätyhjällä ylhäältä rajoitetulla reaalilukujen joukolla on olemassa supremum.

Vastaavasti jokaisella epätyhjällä alhaalta rajoitetulla reaalilukujoukolla on olemassa infimum.

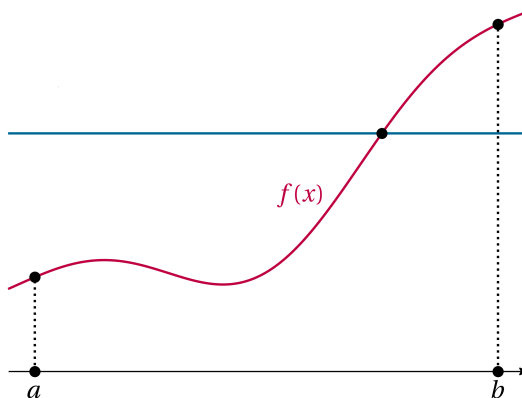
Tämä on itse asiassa niin keskeinen ominaisuus reaaliluvuista, että se otetaan usein osaksi reaalilukujen määritelmää. Se on siis tyypillisesti fakta jota ei todisteta, niin sanottu **aksiooma**, joihin kaikki muu matematiikka perustuu. Tämä kuitenkin riippuu reaalilukujen määrittelytavasta, sillä reaaliluvut voidaan määritellä monella yhtäpitävällä menetelmällä. Tavasta riippuen ylläoleva tulos supremumista saatetaan myös todistaa. Otamme sen tässä aksioomaksi.

Käsittelemme seuraavaksi keskeisen perustuloksen jatkuvuudesta, joka sanoo seuraavaa.

Jatkuvien funktioiden väliarvolause

Olkoon f jatkuva funktio välillä $[a, b]$. Tällöin se saa välin $[a, b]$ sisällä jokaisen arvon lukujen $f(a)$ ja $f(b)$ välissä.

Toisin sanoen: jos $f(a) < u < f(b)$ tai $f(b) > u > f(a)$, niin jollakin välin $[a, b]$ pisteellä c pätee $f(c) = u$.



Kuva 1. Jatkuvien funktioiden väliarvolause.

Lause siis sanoo, että jatkuva funktio ei voi hypätä välillä $[a, b]$ minkään päätepisteiden arvojen välissä olevan arvon yli. Geometrisesti intuitiivinen tapa ymmärtää tämä on, että jatkuvan funktion kuvaaja ei voi hypätä minkään vaakasuoran suoran yli.

Tämän lauseen tärkeä erikoistapaus on niin sanottu **Bolzanon lause**, joka sanoo seuraavaa.

Bolzanon lause

Jos jatkuva funktio saa erimerkkiset arvot pisteissä a ja b , niin sillä on nollakohta avoimella välillä $]a, b[$.

Käydään nyt läpi jatkuvien funktioiden väliarvolauseen todistusidea tiivistettynä tärkeimpiin yksityiskohtiin.

Todistus. Tehdään vastaoletus, että funktio ei saa jotakin arvoa u arvojen $f(a)$ ja $f(b)$ välissä. Käydään symmetrian vuoksi vain tapaus $f(a) < u < f(b)$ läpi.

Olkoon S niiden pisteiden x joukko, jotka kuuluvat välille $[a, b]$, ja joilla $f(x) < u$.

Tällöin S on epätyhjä, sillä a on tällainen piste, ja S :llä on selvä yläraja b . Siis sillä on myös supremum $\sup S = c$. Tutkimme nyt mahdollisia tapauksia.

Tapaus 1: Jos $f(c) = u$. Tällöin väite on todistettu. Näytämme seuraavaksi, että muut tilanteet johtavat ristiriitaan.

Tapaus 2: Jos $f(c) < u$. Olkoon $\varepsilon > 0$ riittävän pieni niin, että $f(c) + \varepsilon < u$. Jatkuvuuden perusteella löytyy $\delta > 0$ siten, että jos $c < x < c + \delta$, niin $|f(x) - f(c)| \leq \varepsilon$. Mutta tällöin $f(x) \leq f(c) + \varepsilon < u$, joten x kuuluu joukkoon S . Siis c ei olekaan joukon S yläraja, sillä $x > c$, mikä on ristiriita.

Tapaus 3: Jos $f(c) > u$. Etsimme nyt ristiriidan näyttämällä, että tällöin f saa suurempia arvoja kuin u joukossa S . Olkoon $\varepsilon > 0$ niin pieni, että $u < f(c) - \varepsilon$. Olkoon $\delta > 0$ nyt sellainen, että $f(x) \geq f(c) - \varepsilon$, kun $|x - c| \leq \delta$ (jatkuvuuden määritelmän nojalla). Jos joukossa S ei olisi pistettä x , jolle $c - \delta \leq x \leq c$, niin c ei olisi pienin yläraja vaan $c - \delta$ olisi pienempi yläraja. Siis tällainen x löytyy, ja nyt $f(x) \geq f(c) - \varepsilon > u$, mikä on ristiriita, sillä $f(x) < u$ joukossa S .

Tutkitaan seuraavaksi jatkuvien funktioiden ääriarvoja. Tähän liittyy pari helppoa peruskäsitettä. Sanomme, että funktio f on **ylhäältä rajoitettu**, jos löytyy luku M siten, että $f(x) \leq M$ kaikilla pistellä x funktion määrittelyjoukossa. Vastaavasti f on **alhaalta rajoitettu**, jos $f(x) \geq m$ jollain m ja kaikilla x . Lukuja M ja m kutsutaan näissä tapauksissa vastaavasti funktion ylä- ja alarajoiksi.

Sanotaan lisäksi, että funktio saavuttaa jossain pisteessä x **suurimman arvonsa** eli **maksimin**, jos $f(x) \geq f(y)$ kaikilla pisteillä y funktion määrittelyjoukossa, eli jos $f(x)$ on funktion yläraja. Vastaavasti funktio saavuttaa **pienimmän arvonsa** eli **minimin**, jos se saavuttaa jossakin pisteessä oman alarajansa. Näitä arvoja kutsutaan funktion **ääriarvoiksi**.

Jatkuvat funktiot ovat mukavia, sillä suljetulla välillä niillä on aina suurin ja pienin arvo.

Weierstrassin lause (jatkuvien funktioiden ääriarvolause)

Mikäli funktio f on jatkuva välillä $[a, b]$, sillä on välillä $[a, b]$ suurin ja pienin arvo.

Todistus. Osoitetaan suurimman arvon tapaus, ja pienimmän arvon tapaus menee samalla tavalla. Osoitetaan ensin, että f on ylhäältä rajoitettu välillä $[a, b]$.

Jos funktio f ei olisi ylhäältä rajoitettu, niin jokaisella positiivisella kokonaisluvulla löytyisi piste x_n väliltä $[a, b]$ siten, että $f(x_n) > n$.

Tarvitsemme nyt perustuloksen reaali-lukujonoista, jonka todistuksen jätämme kuitenkin välistä. Tulos sanoo, että ylhäältä ja alhaalta rajoitetulla jonolla on aina osajono joka suppenee. Eli pisteistä $x_1, x_2, \dots$ voidaan jättää osa pois niin, että

tuotetaan uusi jono $y_1, y_2, \dots$, jolla on olemassa raja-arvo $y = \lim_{n \rightarrow \infty} y_n$. Tulos ei ole hirveän vaikea todistaa, mutta hieman työläs. Se vaatii, että ensin jonosta (x_n) löydetään kasvava tai laskeva osajono, ja sitten perustelun, että kasvava/laskeva jono joka on ylhäältä/alhaalta rajoitettu suppenee aina.

Siispä jonolle (x_n) löytyy osajono (y_n) , joka suppenee johonkin pisteeseen y . Piste y on edelleen välillä $[a, b]$ (jätämme harjoitustehtäväksi), joten voimme nyt jatkuvuuden perusteella päätellä, että jono $f(y_n)$ suppenee kohti arvoa $f(y)$. Mutta jono $f(y_n)$ kasvaa rajatta, sillä (y_n) oli osajono jonosta (x_n) , jolle $f(x_n) > n$. Tämä ristiriita näyttää vihdoin, että funktio f on ylhäältä rajoitettu välillä $[a, b]$.

Koska f on ylhäältä rajoitettu, on sen arvoilla $f(x)$ olemassa supremum M välillä $[a, b]$. Väitämme, että supremum M on itse asiassa maksimi, eli että se saavutetaan jossain pisteessä. Jokaisella n voimme nimittäin löytää pisteen x_n , jolle $f(x_n) > M - \frac{1}{n}$, sillä muuten $M - \frac{1}{n}$ olisi pienempi alaraja kuin M . Etsitään jonolle (x_n) taas osajono (y_n) joka suppenee. Tällöin

$$M \geq f(y) = \lim_{n \rightarrow \infty} f(y_n).$$

Toisaalta, jos $f(y_n) \geq M - \frac{1}{n}$ kaikilla n , niin väistämättä $\lim_{n \rightarrow \infty} f(y_n) \geq M$. Siispä $M \geq f(y) \geq M$, joten $f(y) = M$. Väite on todistettu, sillä f saavuttaa maksiminsa pisteessä y . $\square$

3.3 Derivoituvuus

Määritellään seuraavaksi, mitä funktion derivoituvuus tarkoittaa. Derivaatan määritelmä sisältyy lukion oppimäärään: Derivaatta määritellään **erotusosamäärän** raja-arvona.

Derivaatan määritelmä

Sanotaan, että avoimella välillä $]a, b[$ määritelty funktio f on **derivoituva** tämän välin pisteessä x , jos erotusosamäärällä

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

on raja-arvo, kun $h \rightarrow 0$. Tätä raja-arvoa merkitään $f'(x)$ tai $\frac{df}{dx}$, ja sitä kutsutaan funktion f **derivaataksi** kohdassa x .

Tutkimme tässä siis funktion erotusosamäärää $D(h) = \frac{f(x+h) - f(x)}{h}$ muuttujan h funktiona, ja puhumme sen raja-arvosta kun $h \rightarrow 0$ käyttäen aiemmin esittelemäämme funktion raja-arvon määritelmää. Tässä erityisesti piste 0 ei kuulu erotusosamäärän $D(h)$ määrittelyjoukkoon, mutta vaadimme derivoituvuudessa silti, että sillä on raja-arvo tässä pisteessä.

Derivaatan määritelmästä lähtien voidaan todistaa erilaisille funktioille derivointikaavat, kuten $\frac{d}{dx}(x^3) = 3x^2$. Keskitymme seuraavaksi tiettyihin keskeisiin tuloksiin derivoituvista funktioista.

Fermat'n lause ääriarvoista

Olkoon funktio f jatkuva välillä $[a, b]$ ja derivoituva välillä $]a, b[$. Jos f saavuttaa suurimman tai pienimmän arvonsa avoimen välin $]a, b[$ pisteessä c , niin $f'(c) = 0$.

Todistus. Käsitellään tapaus, jossa $f(c)$ on funktion maksimi. Erotusosamäärää tutkimalla havaitaan, että jos $h > 0$, niin

$$\frac{f(c+h) - f(c)}{h} \leq \frac{f(c) - f(c)}{h} = 0.$$

Toisaalta

$$\frac{f(c-h) - f(c)}{-h} \geq \frac{f(c) - f(c)}{-h} = 0.$$

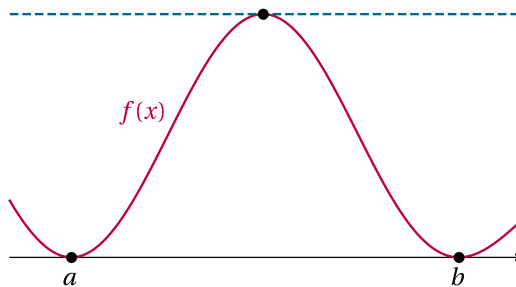
Epäyhtälöt säilyvät, kun otetaan raja-arvo kun $h \rightarrow 0$, joten $f'(c) = 0$. □

Fermat'n lause sanoo erityisesti, että derivoituva funktio saavuttaa suljetulla välillä ääriarvonsa joko välin päätepisteissä tai derivaatan nollakohdassa. Siitä voidaan myös johtaa seuraava tärkeä aputulos.

Rollen lause

Jos funktio f on jatkuva välillä $[a, b]$, derivoituva välillä $]a, b[$, ja $f(a) = f(b) = 0$, niin väliltä $]a, b[$ löytyy piste ξ , jolle $f'(\xi) = 0$.

Tässä käytetty kirjain ξ on kreikkalainen *ksii*.



Kuva 2. Rollen lause.

Todistus. Koska f on jatkuva, se saavuttaa ääriarvonsa välillä $[a, b]$. Jos sekä maksimi että minimi saavutetaan päätepisteissä, niin molemmat ovat oletuksen perusteella arvoltaan 0. Tällöin $f(x) = 0$ kaikissa välin pisteissä, jolloin selvästi väite on totta.

Jäljelle jäävässä tapauksessa funktio f saavuttaa ainakin yhden ääriarvoistaan avoimella välillä $]a, b[$. Tällöin väite seuraa edellisestä lauseesta. $\square$

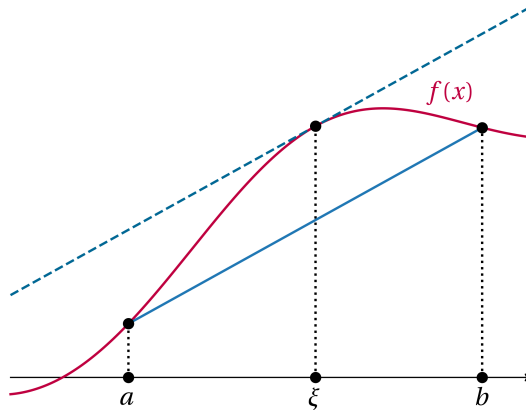
Käydään vielä läpi hyvin keskeinen derivoituvuuteen liittyvä lause, jota tulemme soveltamaan monella tapaa tulevilla luvuilla.

Differentiaalilaskennan väliarvolause

Jos funktio f on jatkuva välillä $[a, b]$ ja derivoituva välillä $]a, b[$, niin väliltä $]a, b[$ löytyy piste ξ , jolle pätee

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

Sanomme jatkossa differentiaalilaskennan väliarvolauseetta vain **väliarvolauseeksi**. Huomataan, että väliarvolause on suora yleistys Rollen lauseesta, ja sen todistus perustuukin Rollen lauseeseen.



Kuva 3. Väliarvolause takaa sopivan tangentin olemassaolon.

Todistus. Väite seuraa suoraan Rollen lauseesta sovellettuna uuteen funktioon

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Nimittäin helposti nähdään, että $g(a) = g(b) = 0$, ja Rollen lauseen nojalla seuraa, että jollakin pisteellä $\xi \in]a, b[$ on $g'(\xi) = 0$. Laskemalla funktion g derivaatta

saadaan

$$0 = g'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{b - a},$$

mistä siirtämällä termit eri puolille saadaan haluttu kaava (2). $\square$

Geometrisesti väliarvolause sanoo, että jos välin $[a, b]$ päätepisteisiin sovittaa suoran, joka kulkee funktion kuvaajan pisteiden $(a, f(a))$ ja $(b, f(b))$ kautta, niin funktiolla täytyy tällä välillä olla jossain pisteessä ξ tangentti, jonka kulmakerroin $f'(\xi)$ on sama kuin tällä suoralla.

Harjoitustehtäviä

Tehtävä 22. Osoita raja-arvon määritelmästä, että jonon $x_n = \frac{2}{n}$ raja arvo on 0.

Vihje s. 156 $\rightarrow$

Toinen vihje s. 170 $\rightarrow$

Tehtävä 23. Osoita, että funktion $f(x) = \frac{1}{x-1}$ raja-arvo kohdassa $x = 2$ on 1.

Vihje s. 156 $\rightarrow$

Toinen vihje s. 170 $\rightarrow$

Tehtävä 24. Osoita raja-arvon määritelmästä lähtien, että jonolla $x_n = (-1)^n$ ei ole raja-arvoa.

Vihje s. 157 $\rightarrow$

Toinen vihje s. 170 $\rightarrow$

Tehtävä 25. Olkoon $f : [0, 1] \rightarrow [0, 1]$ jatkuva kuvaus, eli f on määritelty ja saa arvoja välillä $[0, 1]$. Osoita, että funktiolla f on **kiintopiste** x , eli että yhtälöllä $f(x) = x$ on ratkaisu välillä $[0, 1]$.

Vihje s. 157 $\rightarrow$

Toinen vihje s. 170 $\rightarrow$

Tehtävä 26. Osoita raja-arvon määritelmän avulla jatkuvien funktioiden ääriarvolauseessa käytetty aputuloks: Jos jonolla (y_n) on olemassa raja-arvo y , ja jokainen jonon jäsen y_n on luku väliltä $[a, b]$, niin myös y on luku väliltä $[a, b]$.

Vihje s. 157 $\rightarrow$

Toinen vihje s. 171 $\rightarrow$

Tehtävä 27. Osoita väliarvolauseeseen avulla, että jos f on jatkuva, derivoituva, ja $f'(x) = 0$ kaikilla reaaliluvuilla x , niin f on vakiofunktio.

Vihje s. 157 $\rightarrow$

Toinen vihje s. 171 $\rightarrow$

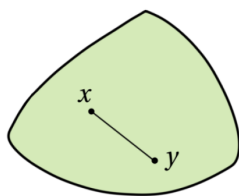
Tehtävä 28. Osoita, että jos f on jatkuva, derivoituva, ja $f'(x) > 0$ kaikilla x , niin f on aidosti kasvava funktio.

Vihje s. 157 $\rightarrow$

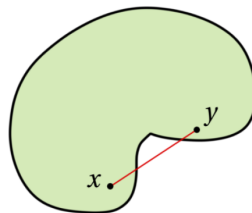
Toinen vihje s. 171 $\rightarrow$

4 Konveksit funktiot

Konveksisuus (eng. *convex*) on perustavanlaatuinen käsite matematiikassa, jolla on monipuolisia sovelluksia eri matematiikan osa-alueilla. Kontekstista riippuen konveksisuus voi tarkoittaa hieman eri asioita. Geometrisen kappaleen konveksisuus tarkoittaa sitä, että jos kappaleesta valitaan mitkä tahansa kaksi pistettä, niin niiden välille piirretty jana on edelleen kappaleen sisällä. Esimerkiksi ellipsit ja suorakulmiot ovat konvekseja kappaleita.



Konvekxi kappale



Epäkonvekxi kappale

Geometrista konveksisuutta voidaan täsmentää käyttämällä vektorinotaatiota: Olkoon S geometrinen kappale, eli esimerkiksi tason $\mathbb{R}^2$ tai kolmiulotteisen avaruuden $\mathbb{R}^3$ osajoukko. Tällöin S on konvekxi, jos jokaisella kappaleeseen S kuuluvalla pisteparilla x, y ja vakiolla λ , missä $0 \leq \lambda \leq 1$, myös piste

$$(1 - \lambda)x + \lambda y$$

kuuluu joukkoon S . Tämä lauseke nimittäin esittää jotakin pistettä pisteiden x ja y välisellä janalla, missä parametri λ kertoo mistä kohtaa janaa piste on valittu (esim. parametrilla $\lambda = 1/2$ saadaan janan keskipiste $(x + y)/2$).

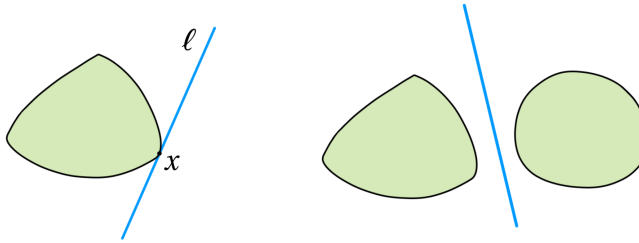
Kappaleen konveksisuudella on myös tärkeä sidos sen reunan kaarevuuteen. Kaksiulotteinen kappale on konvekxi täsmälleen silloin, kun sen jokaista reunapistettä sivuaa suora siten, että kappale jää kokonaan suoran toiselle puolelle (suora mukaan lukien). Sama pätee kolmessa ulottuvuudessa, kun suora korvataan tasolla. Lyhyesti sanottuna tämä tarkoittaa, että konveksin kappaleen reuna on kupera.

Tähän liittyy myös perustavanlaatuinen tulos kahden konveksin kappaleen erottelusta, nimeltä **erottavan hypertason lause**. Kahdessa ulottuvuudessa lause sanoo seuraavaa.

Lause (Erottavan hypertason lause tasossa)

Jos kaksi tason konveksia kappaletta eivät leikkaa, niiden välille voidaan piirtää sellainen suora, että molemmat kappaleet ovat suoran eri puolilla.

Suoran koskettaminen sallitaan tässä myös. Kolmessa ulottuvuudessa suoran voi taas korvata tasolla, ja lause pätee vastaavalla tavalla myös useampiulotteisille kappaleille. Hahmottelemme todistusidean lyhyesti.



Kuva 4. Konveksien kappaleiden erityisominaisuuksia.

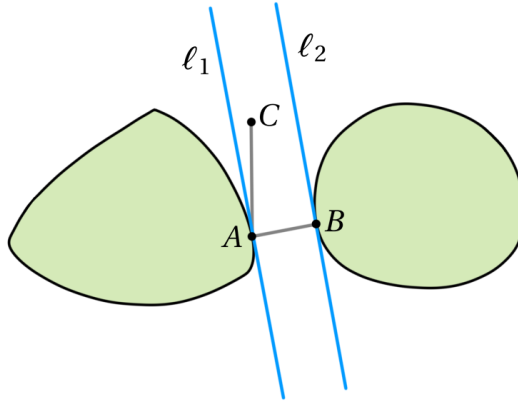
Todistus. Olkoon S ja T kaksi tason konveksia kappaletta, jotka eivät leikkaa toisiaan. Oletetaan tässä, että molemmat kappaleet ovat rajoitettuja (eivät sisällä pisteitä mielivaltaisen kaukana toisistaan, esim. puolitaso ei ole rajoitettu) ja suljettuja (kappaleet sisältävät oman reunansa). Ensimmäinen idea todistuksessa on tutkia, mikä on lyhin etäisyys, joka kappaleiden S ja T pisteiden väliltä voidaan löytää. On mahdollista osoittaa, että löytyy sellaiset pisteet A ja B kappaleiden S ja T reunoilta, jotka minimoivat tämän etäisyyden. Pisteiden A ja B etäisyyttä sanotaan myös kappaleiden S ja T etäisyydeksi.

Piirretään nyt pisteisiin A ja B kaksi suoraa ℓ_1 ja ℓ_2 , jotka ovat kohtisuorassa janaa AB vasten. Väitämme, että näiden yhdensuuntaisten suorien ℓ_1 ja ℓ_2 väliltä ei löydy yhtään kappaleiden S ja T pistettä. Jos esimerkiksi tältä väliltä löytyisi kappaleen S piste C , niin myös koko janan AC täytyisi sisältyä kappaleeseen S konveksisuuden perusteella. Mutta koska janan AC muodostama kulma janan AB kanssa on terävä, voidaan tasogeometrialla osoittaa, että janalta AC löytyy pisteitä, jotka ovat lähempänä pistettä B kuin piste A . Tämä on ristiriita, sillä A ja B olivat sellaiset pisteet, jotka minimoivat kappaleiden S ja T etäisyyden, joten suorien ℓ_1 ja ℓ_2 välillä ei voi olla kappaleiden pisteitä. Tällöin mikä tahansa janaa AB leikkaava kohtisuora suora erottelee kappaleet S ja T puolilleen.

Lause pätee myös silloin, kun kappaleet S ja T ovat rajoittamattomia. Esim. paraabeli rajaa tasosta rajoittamattoman konveksin alueen. Tässä tapauksessa todistus

vaatii lisäksi rajankäyntiargumentin, jossa rajoittamatonta aluetta lähestytään rajoitetuilla konvekseilla kappaleilla.

Tässä esitetty kaksiulotteinen todistusidea toimii samaan tyyliin myös useammas-
sa ulottuvuudessa. □



Kuva 5. Erottavan hypertason lauseen todistusidea.

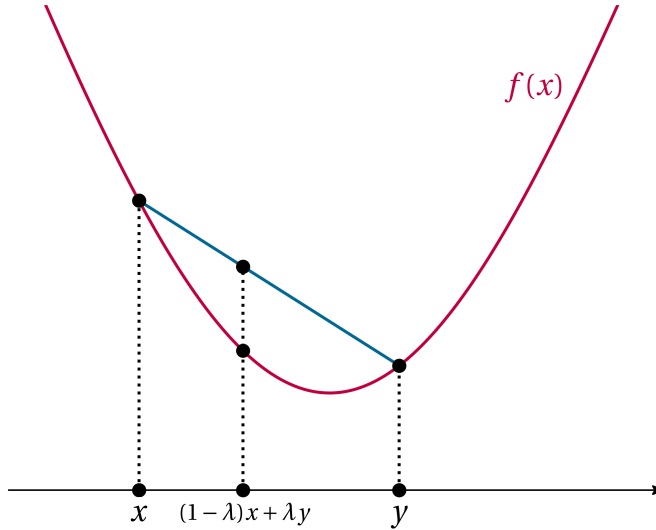
Erottavan hypertason lauseella on syvällisiä sovelluksia erityisesti optimointiongelmassa. Yksi tyypillinen sovelluskohde on tietokonepelien fysiikkasimulaatioissa, jossa lause on tärkeässä roolissa kun rakennetaan algoritmeja havaitsemaan, milloin kaksi kappaletta törmäävät.

Erittelemme jatkossa geometrisen kappaleen konveksisuuden termillä **geometrisen konveksisuus**.

4.1 Funktion konveksisuus

Konveksisuudella on uusi merkitys funktioista puhuttaessa. Olkoon $f : [a, b] \rightarrow \mathbb{R}$ funktio, joka on määritelty reaaliakselin suljetulla välillä $[a, b]$. Sanomme, että funktio f on **konvekksi** eli **alaspäin kupera**, jos jokaisella pisteparilla x, y väliltä $[a, b]$ pätee seuraava geometrinen ominaisuus: Kun funktion kuvaajasta valitsee näitä koordinaatteja vastaavat pisteet $(x, f(x))$ ja $(y, f(y))$, niin niiden välille piirretty jana ei missään kohdassa mene funktion kuvaajan alle. Tämä ominaisuus on esitetty Kuvassa 6, ja se on yhtäpitävää sen kanssa, että funktion kuvaajan ylle jäävä alue on geometrisesti konvekksi.

Funktion konveksisuus voidaan ilmaista myös epäyhtälönä. Olkoot taas x ja y pisteitä funktion f määrittelyväliltä, ja olkoon λ mikä tahansa parametri väliltä $0 \leq \lambda \leq 1$. Tällöin piste $a = (1 - \lambda)x + \lambda y$ esittää mielivaltaista koordinaattia pisteiden x ja y välistä. Funktion kuvaajan korkeuden tässä pisteessä a kertoo sen arvo



Kuva 6. Konveksin funktion määritelmä geometrisesti.

$f(a)$, kun taas janalla pisteiden $(x, f(x))$ ja $(y, f(y))$ välillä vastaavassa kohdassa pisteen korkeus saadaan kaavasta $(1-\lambda)f(x) + \lambda f(y)$. Funktion f konveksisuuden perusteella tämä jälkimmäinen korkeus on suurempi, joten saadaan epäyhtälö

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y). \quad (3)$$

Tämän epäyhtälön voimassaolo kaikilla sallituilla parametrien x, y , ja λ valinnoilla on myös yhtäpitävää funktion f konveksisuuden kanssa. Siis epäyhtälö (3) luokittelee kaikki konveksit funktiot.

Vaikka konveksin funktion määrittelevä epäyhtälö (3) on mahdollista ymmärtää myös silloin kun funktio f ei ole jatkuva, osoittautuu, että jokainen konvekssi funktio on kuitenkin jatkuva.

4.2 Konveksisuus ja derivaatat

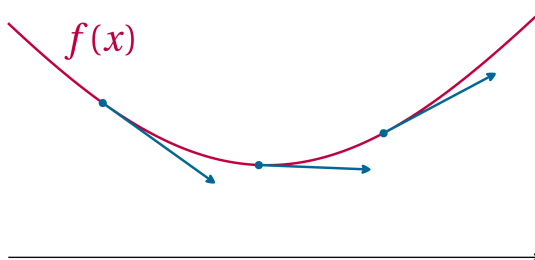
Konveksin funktion ei tarvitse olla derivoituva kaikkialla. Esimerkiksi itseisarvo-funktio $f(x) = |x|$ on konvekssi, mutta sillä ei ole derivaattaa pisteessä $x = 0$. Jos kuitenkin funktiomme on derivoituva, on konveksisuudella tietty yhteys funktion derivaatan käyttäytymiseen.

Derivoituvuus kertoo ensinnäkin, että funktion f kuvaajalle voidaan piirtää joka pisteessä x tangenttisuora, jonka kulmakertoimen ilmaisee derivaatta $f'(x)$. Osoittautuu, että tämä tangenttisuora antaa meille kaksi uutta tapaa luokitella funktion konveksisuus.

Konveksien funktioiden perusominaisuus 1

Jatkuvasti derivoituva funktio on konvekksi täsmälleen silloin, kun sen derivaatta on kasvava funktio.

Lause väittää siis, että konveksin funktion kuvaajalle piirrettyjen tangenttisuorien derivaatat kasvavat kuvaajalla edetessä. Tätä tulosta on havainnollistettu geometrisesti oheisessa kuvassa.



Kuva 7. Konveksin funktion derivaatta, eli tangentin kulmakerroin, kasvaa kuvaajalla edetessä.

Derivaatan kasvavuuden tarkka perustelu onnistuu pyörittämällä funktion erotusosamäärää ja kaavaa 3, mutta säästämme tässä tilaa ja jätämme perustelun geometrisen intuition varaan. Toinen tangenttisuoriin liittyvä ehto on seuraava.

Konveksien funktioiden perusominaisuus 2

Jatkuvasti derivoituva funktio on konvekksi täsmälleen silloin, kun sen jokainen tangenttisuora jää kuvaajan alapuolelle.

Tämän lauseen sanoman voi myös esittää epäyhtälönä: Pisteessä x_0 kuvaajan pisteen $(x_0, f(x_0))$ kautta kulkevan tangenttisuoran yhtälö on $y = f'(x_0)(x - x_0) + f(x_0)$, joten Lause 2 voidaan ilmaista epäyhtälönä

$$f(x) \geq f'(x_0)(x - x_0) + f(x_0), \quad (4)$$

joka pätee kaikilla muuttujan x arvoilla. Erityisesti tässä nähdään, että konvekssi funktio kasvaa joka pisteestä eteenpäin ainakin yhtä nopeasti kuin sille piirretty tangenttisuora. Myös tämä tulos on geometrisesti varsin uskottava, joten jätämme yksityiskohtaisen todistuksen pois.

Toinen derivaatta. Mikäli funktio f on kahdesti derivoituva, eli sille löytyy toinen derivaatta f'' , voidaan derivaatan f' kasvavuus ilmaista toisen derivaatan f'' ei-negatiivisuutena. Tästä saamme jälleen kerran uuden ehdon konveksisuudelle.

Konveksien funktioiden perusominaisuus 3

Kahdesti derivoituva funktio f on konvekxi täsmälleen silloin, kun sen toinen derivaatta f'' on ei-negatiivinen, eli $f''(x) \geq 0$ kaikilla x .

Siispä siinä missä funktion ensimmäisen derivaatan rooli on kertoa funktion kasvavuudesta, on toisen derivaatan rooli luokitella funktion konveksisuus. Toisen derivaatan laskeminen on monessa tapauksessa myös paljon yksinkertaisempaa kuin epäyhtälöehdon (3) tarkistaminen jokaisella pisteparilla, joten tämä ominaisuus tarjoaa myös käytännöllisen tavan tarkistaa, onko annettu funktio konvekxi.

Lisää konveksisuudesta. Konveksisuudelle on myös käänteinen ehto nimeltä **konkaavius** (eng. concave). Funktio on konkaavi tai **ylöspäin kupera**, jos se toteuttaa epäyhtälön (3) käänteisen version

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y). \quad (5)$$

Funktio f on konkaavi täsmälleen silloin, kun funktio $-f$ on konvekxi, ja päinvastoin. Erityisesti esimerkiksi funktion toisen derivaatan negatiivisuus kertoo, että funktio on konkaavi.

Lisäksi sanomme, että funktio f on määrittelyvälillään **aidosti konvekxi**, jos se toteuttaa epäyhtälön (3) aidon version, eli kun

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y) \quad (6)$$

kaikilla pisteillä x, y ja parametreilla λ , missä $0 < \lambda < 1$. Tämä tarkoittaa yhtäpitävästi sitä, että kuvaajan pisteiden väliin piirretty jana makaa aidosti funktion kuvaajan yläpuolella päätepisteitä lukuunottamatta. Vastaavalla tavalla määritellään myös **aidosti konkaavit** funktiot.

Tiedetään myös, että derivoituva funktio on aidosti konvekxi täsmälleen silloin, kun sen derivaatta on aidosti kasvava. Tämä ei kuitenkaan tarkoita, etteikö toisella derivaatalla voisi silti olla yksittäisiä nollakohtia, kuten aidosti konveksilla funktiolla $f(x) = x^4$.

4.3 Jensenin epäyhtälö

Konveksien funktioiden eräs tärkeä sovellus on **Jensenin epäyhtälö**. Olkoon funktio f konvekxi, ja painot $\lambda_i \geq 0$, missä $i = 1, \dots, n$, sellaisia, että $\sum_{i=1}^n \lambda_i = 1$. Tällöin

mille tahansa pisteille $x_1, \dots, x_n$ funktion f määrittelyjoukosta pätee epäyhtälö

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i), \quad (7)$$

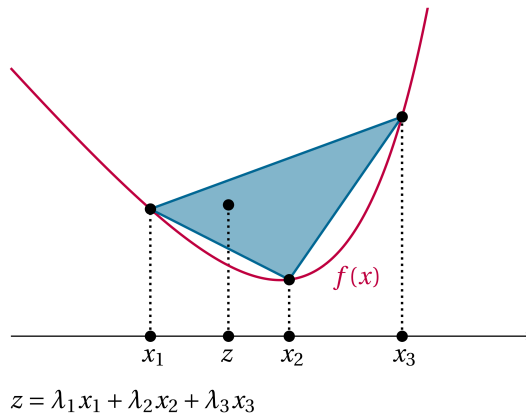
jota kutsutaan klassiseksi Jensenin epäyhtälöksi.

Jensenin epäyhtälö on varsin suoraviivaista todistaa induktiolla suoraan konveksisuuden määritelmästä (3), joka on sama kun Jensenin epäyhtälö kahden pisteen tapauksessa. Emme käy tätä induktiotodistusta tässä läpi, vaan esitämme sen sijaan luontevan geometrisen perustelun epäyhtälölle.

Tutkitaan epäyhtälöä ensin kolmen pisteen tapauksessa, jossa se sanoo, että konveksille funktiolle pätee

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3), \quad (8)$$

missä $\lambda_1 + \lambda_2 + \lambda_3 = 1$ ja $\lambda_1, \lambda_2, \lambda_3 \geq 0$.



Kuva 8. Jensenin epäyhtälö kertoo oleellisesti sen, että konveksin funktion käyrän pisteisiin piirretty monikulmio makaa sen kuvaajan yläpuolella.

Tutkitaan funktion kuvaajan kolmea pistettä $(x_1, f(x_1))$, $(x_2, f(x_2))$, ja $(x_3, f(x_3))$. Nämä kolme pistettä rajaavat tasoon kolmion, ja osoittautuu, että jokaisen tämän kolmion pisteen voi esittää koordinaattimuodossa

$$\begin{aligned} & \lambda_1(x_1, f(x_1)) + \lambda_2(x_2, f(x_2)) + \lambda_3(x_3, f(x_3)) \\ &= (\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3, \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3)), \end{aligned}$$

missä painokertoimet λ_i ovat edelleen ei-negatiivisia ja toteuttavat ehdon $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Yleinen kolmion piste saadaan siis painottamalla sopivasti sen kärkipisteitä.

Vertaamalla Jensenin epäyhtälöön (8) huomataan, että epäyhtälö oikeastaan väittää sitä, että funktion arvo pisteessä $\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3$ on pienempi kuin vastaavan

kolmion pisteen y -koordinaatti. Tiivistäen, Jensenin epäyhtälö kolmen pisteen tapauksessa siis sanoo, että konveksin funktion kuvaajan pisteisiin piirretty kolmio makaa sen kuvaajan yläpuolella.

Yleisessä tapauksessa Jensenin epäyhtälö sanoo siis geometrisesti sitä, että konveksin funktion kuvaajan pisteisiin piirretty monikulmio makaa sen kuvaajan yläpuolella. Tämä on varsin uskottavaa, sillä monikulmion alareuna koostuu janoista, jotka tiedetään jo makaavan kuvaajan yläpuolella konveksisuuden määritelmän perusteella.

Jensenin epäyhtälön integraaliversio. Jensenin epäyhtälöstä pätee myös vielä hieman yleisempi versio, jossa summat korvataan määrättyllä integraalilla. Olkoon funktio f edelleen konvekksi, ja g mikä tahansa jatkuva funktio joka saa ei-negatiivisia arvoja. Tällöin pätee seuraava epäyhtälö

$$f\left(\frac{1}{b-a}\int_a^b g(x)dx\right) \leq \frac{1}{b-a}\int_a^b f(g(x))dx. \quad (9)$$

Jensenin epäyhtälön versioilla on paljon erilaisia sovelluksia. Konkaaveille funktioille näiden epäyhtälöiden merkki luonnollisesti kääntyy toisin päin.

Esimerkki. Todistetaan neljän muuttujan aritmeettis-kvadraattinen epäyhtälö

$$\frac{a+b+c+d}{4} \leq \sqrt{\frac{a^2+b^2+c^2+d^2}{4}},$$

missä $a, b, c, d \in \mathbb{R}$. Funktio $f(x) = x^2$ on konvekksi, sillä sen toinen derivaatta $f''(x) = 2$ on positiivinen kaikilla $x \in \mathbb{R}$. Voidaan siis kirjoittaa

$$\left(\frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d\right)^2 \leq \frac{1}{4}a^2 + \frac{1}{4}b^2 + \frac{1}{4}c^2 + \frac{1}{4}d^2, \quad \text{eli}$$

$$\left|\frac{a+b+c+d}{4}\right| \leq \sqrt{\frac{a^2+b^2+c^2+d^2}{4}}.$$

Väitteen epäyhtälö seuraa tästä, sillä luku on korkeintaan itseisarvonsa suuruinen. Todistus etenisi täsmälleen samoin, vaikka muuttujia olisi enemmän.

4.4 Optimointiongelmat ja konveksisuus

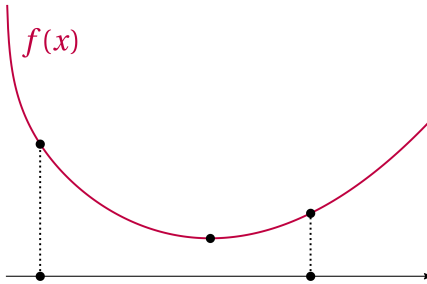
Ehkä tärkein sovellus funktion konveksisuuden käsitteelle on sen yhteys **optimointiongelmiin** kanssa. Tyypillisessä optimointiongelmassa meillä on annettu jokin reaaliarvoinen funktio f , ja tarkoitus on selvittää onko funktiolla f olemassa pienintä (tai suurinta) arvoa määrittelyjoukossaan. Sanotaan, että funktiolla on pienin arvo $f(a)$ pisteessä a , jos $f(x) \geq f(a)$ kaikilla mahdollisilla pisteillä x . Tässä

tapauksessa pistettä a kutsutaan funktion f **minimikohdaksi** tai **minimipisteeksi**. Jos funktio saavuttaa pienimmän arvonsa, on myös luonnollista kysyä, onko sillä yksikäsitteinen minimikohta vai onko niitä useampia. Konveksisuus liittyy läheisesti tähän kysymykseen, sillä se tarjoaa siihen suoran vastauksen:

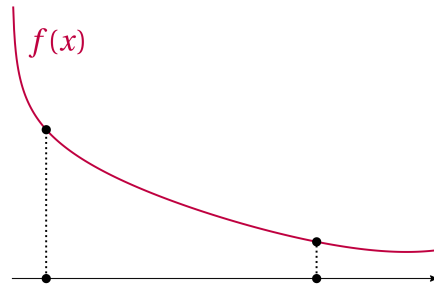
Konveksien funktioiden perusominaisuus 4

Aidosti konveksilla funktiolla on jokaisella suljetulla välillä $[a, b]$ olemassa yksikäsitteinen minimikohta.

Tämä tulos on geometrisesti varsin uskottava, mutta käydään sille läpi myös tarkka matemaattinen perustelu. Jatkuva funktio saavuttaa suljetulla välillä aina pienimmän ja suurimman arvonsa (**Weierstrassin lause**), joten pienimmän arvon olemassaolo on taattu konveksien funktioiden jatkuvuuden nojalla. Siis riittää selvittää, miksi pienintä arvoa ei voi saavuttaa useammassa pisteessä.



Kuva 9. Aidosti konveksin funktion yksikäsitteinen minimikohta suljetulla välillä löytyy aina.



Kuva 10. On myös mahdollista, että minimikohta löytyy välin päätepisteestä.

Tutkitaan mitä tapahtuu, jos funktiolla f olisikin kaksi minimikohtaa x_1 ja x_2 . Sijoittamalla nämä minimikohdat ja parametri $\lambda = \frac{1}{2}$ epäyhtälöön (6) saadaan

$$f\left(\frac{x_1 + x_2}{2}\right) < \frac{f(x_1) + f(x_2)}{2}.$$

Tämä on ristiriita, sillä epäyhtälö osoittaa funktion saavan pisteessä $\frac{x_1 + x_2}{2}$ pienemmän arvon kuin minimikohdissaan. Siis minimikohtia voi olla enintään yksi, joten lause on todistettu.

Mikäli funktiomme on lisäksi derivoituva, tiedetään, että se voi saavuttaa pienimmän arvonsa ainoastaan derivaatan nollakohdissa tai välin päätepisteissä. Aidosti konveksin funktion tapauksessa derivaatta on aidosti kasvava, joten sillä voi olla ainoastaan yksi nollakohta joka löytyessään paljastaa minimikohdan. Mikäli nollakohtaa ei löydy, on funktiomme joko aidosti kasvava tai vähenevä, eli minimi saavutetaan toisessa päätepisteistä.

4.5 Usean muuttujan analyysiä

Funktioiden konveksisuus on laajasti sovellettava käsite, sillä se voidaan määritellä myös usean muuttujan funktioille. Erityisesti optimointiongelmassa esiintyy usein tilanteita, joissa optimoitava suure voi riippua lukuisista eri muuttujista. Käsittelemme tässä selkeyden vuoksi kahden muuttujan funktioita $f(x, y)$, mutta teoria yleistyy samalla tavalla myös useammankin muuttujan tapaukseen.

Kahden muuttujan reaalfunktio $f(x, y)$ on suure, joka riippuu kahdesta reaali-muuttujasta x ja y , ja saa arvoja reaalityyppisissä. Tällaisen funktion kuvaaja voidaan hahmottaa pintana kolmessa ulottuvuudessa, ja kuvaajan pisteet ovat muotoa $(x, y, f(x, y))$. Esimerkiksi lineaarisen kahden muuttujan funktion $f(x, y) = ax + by$ kuvaaja on origon kautta kulkeva taso.

Kahden muuttujan funktion konveksisuus voidaan määritellä täysin samalla tavalla kuin yhden muuttujan tapauksessa, eli funktio on konvekksi, jos sen kuvaajan pisteiden välille piirretty jana ei mene kuvaajan alapuolelle. Tämä on edelleen sama asia kuin se, että funktion kuvaajan ylle jäävä alue on geometrisesti konvekksi.

Epäyhtälölle (3) on myös luonnollinen tulkinta tässä tapauksessa. Kahden reaali-muuttujan funktiota $f(x, y)$ voi tulkita myös yhden vektoripisteen $\mathbf{p} = (x, y)$ funktiona $f(\mathbf{p})$. Tällöin funktion f konveksisuus voidaan ilmaista epäyhtälönä

$$f((1 - \lambda)\mathbf{p} + \lambda\mathbf{q}) \leq (1 - \lambda)f(\mathbf{p}) + \lambda f(\mathbf{q}),$$

missä pisteet $\mathbf{p} = (x, y)$ ja $\mathbf{q} = (x', y')$ ovat mielivaltaisia, ja vasemmalla puolella esiintyvä piste tulkitaan $(1 - \lambda)\mathbf{p} + \lambda\mathbf{q} = ((1 - \lambda)x + \lambda x', (1 - \lambda)y + \lambda y')$. Vastaavasti myös aito konveksisuus ja konkaavitus pystytään määrittelemään usean muuttujan funktioille samalla tavalla kuin aiemminkin.

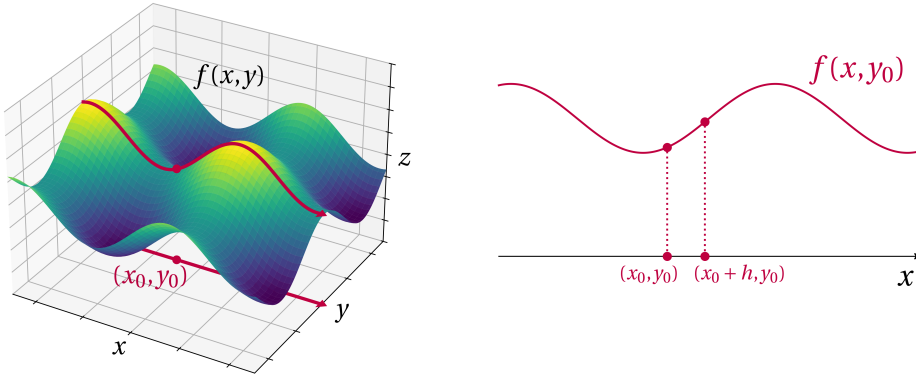
Tutkitaan seuraavaksi, miten useamman muuttujan funktioita derivoidaan.

Osittaisderivaatat. Useamman muuttujan funktioista puhuttaessa derivaatan käsite saa uuden merkityksen. Kahden muuttujan tapauksessa ei ole mielekäästä pyrkiä määrittelemään derivaattaa funktion kasvunopeutena, sillä kasvunopeuden käsite on epämääräinen ellei ensin sovita, mistä suunnasta puhutaan. Kun funktiomme $f(x, y)$ riippuu kahdesta eri muuttujasta x ja y , voimme tutkia sen kasvunopeutta x - ja y -suunnissa erikseen niin sanottujen **osittaisderivaattojen** avulla. Näitä osittaisderivaattoja merkitään $\frac{\partial f}{\partial x}$ ja $\frac{\partial f}{\partial y}$, ja ne määritellään erotusosamäärien raja-arvoina

$$\frac{\partial f}{\partial x}(x, y) = \lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x, y)}{h} \quad \text{ja} \quad \frac{\partial f}{\partial y}(x, y) = \lim_{h \rightarrow 0} \frac{f(x, y + h) - f(x, y)}{h}.$$

Esimerkiksi x -suuntainen osittaisderivaatta $\frac{\partial f}{\partial x}$ ilmaisee funktion $f(x, y)$ kasvunopeuden x -suunnassa, ja sen voi laskea yksinkertaisesti ajattelemalla muuttujaa y vakiona ja laskemalla lausekkeen $f(x, y)$ derivaatta muuttujan x suhteen kuten

yhden muuttujan tapauksessa. Esimerkiksi funktion $f(x, y) = yx^2 + 2y$ derivaatta muuttujan x suhteen on $\frac{\partial f}{\partial x} = 2yx + 0 = 2xy$.



Kuva 11. Kahden muuttujan funktion x -suuntaisen osittaisderivaatan $\frac{\partial f}{\partial x}$ voi laskea pisteessä (x_0, y_0) kiinnittämällä y -koordinaatin ja laskemalla muuttujasta x riippuvan funktion $f(x, y_0)$ derivaatan tavalliseen tapaan.

Yhden muuttujan tapauksessa eräs tärkeä johtopäätös derivoituvuudesta on ensimmäisen asteen Taylorin kehittämän olemassaolo, eli mahdollisuus approksimoida derivoituvaa funktiota sen tangenttisuorilla. Myös useamman muuttujan tapauksessa on hyödyllistä ymmärtää, milloin funktiotamme on mahdollista approksimoida lineaarisella kuvauksella. Tähän ei riitä olettaa, että osittaisderivaatat ovat olemassa, vaan vaaditaan vahvempi käsite nimeltään **derivoituvuus**. Kahden muuttujan funktiota $f(x, y)$ sanotaan derivoituvaksi, jos sillä on jokaisen pisteen (x_0, y_0) suhteen olemassa ensimmäisen asteen Taylorin kehitemä

$$f(x, y) = f(x_0, y_0) + A(x - x_0) + B(y - y_0) + R(x, y), \quad (10)$$

missä A ja B ovat vakioita jotka voidaan tarvittaessa laskea osittaisderivaattojen avulla kaavoista

$$A = \frac{\partial f}{\partial x}(x_0, y_0) \quad \text{ja} \quad B = \frac{\partial f}{\partial y}(x_0, y_0),$$

ja jäännöstermi $R(x, y)$ on lähellä nollaa silloin, kun piste (x, y) on lähellä pistettä (x_0, y_0) , eli kun etäisyys $|(x, y) - (x_0, y_0)| = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ on pieni. Funktion derivoituvuus siis yksinkertaisesti kertoo sen, että jokaisen pisteen lähetyillä se käyttäytyy kuten affiini kuvaus

$$L_{(x_0, y_0)}(x, y) = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0). \quad (11)$$

Osittaisderivaatoista on mahdollista muodostaa vektoriarvoinen funktio $\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y) \right)$, jota kutsutaan funktion f **gradientiksi**. Tätä käytetään ilmaisemaan edellä mainitut kaavat lyhyemmässä muodossa, esimerkiksi lineaarikuvauksen $L_{(x_0, y_0)}$ esitetään usein vektoripistetulon avulla muodossa

$$L_{(x_0, y_0)}(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0).$$

Derivoituvan kahden muuttujan funktion kuvaajaa sivuaa joka pisteessä sen **tangenttitaso**. Tämä taso vastaa tismalleen lineaarikuvauksen (11) kuvaajaa, eli tangenttitasolla sijaitsevat pisteet (x, y, z) saadaan yhtälöstä

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0),$$

missä tangenttitaso on muodostettu kuvaajan pisteeseen $(x_0, y_0, f(x_0, y_0))$.

Derivoituva kahden muuttujan funktio on konveksi täsmälleen silloin, kun sen kuvaaja sijaitsee tangenttitason yläpuolella, eli kun pätee

$$f(x, y) \geq f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (12)$$

jokaisella pisteellä (x, y) funktion määrittelyjoukossa.

Usean muuttujan optimointiongelmat

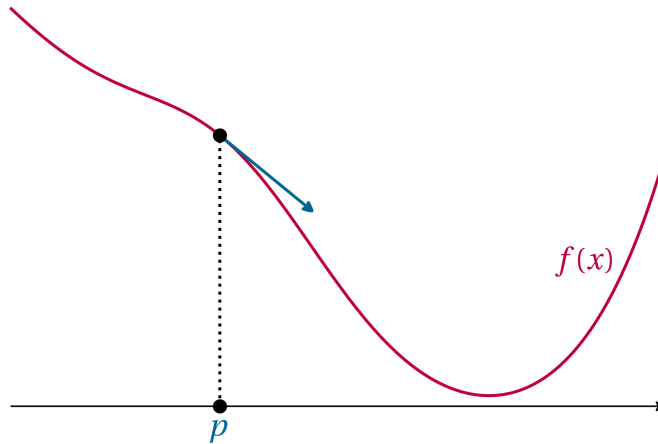
Usean muuttujan funktion minimiarvon löytäminen on hieman haastavampaa kuin yhden muuttujan tapauksessa. Konveksisuus on edelleen hyödyllinen käsite minimiarvon luokittelussa, sillä aidosti konveksilla useamman muuttujan funktiolla on edelleen olemassa yksikäsitteinen minimiarvo (kunhan sen määrittelyjoukko on geometrisesti konveksi, esimerkiksi suorakaide $[a, b] \times [c, d]$ kahden muuttujan tapauksessa).

Mikäli derivoituvan aidosti konveksin funktion gradientilla on nollakohta, eli piste, jossa kaikki funktion osittaisderivaatat ovat nollia, niin tämä piste on myös funktion globaali minimi. Tämä selviää helposti epäyhtälöstä (12), sillä tällaisessa pisteessä lineaariset termit häviävät ja jäljelle jää vain epäyhtälö $f(x, y) \geq f(x_0, y_0)$, joka sanoo, että (x_0, y_0) on funktion minimikohta.

Kuitenkin monessa tapauksessa gradientin nollakohdan määrittäminen on vaikeaa, ja käytännön minimointiongelmissa minimoitava suure ei välttämättä käytäyty kuten konveksi funktio. On kuitenkin muita menetelmiä, joilla funktioiden minimiarvoja voi lähteä etsimään. Tunnetuin näistä on **gradienttimenetelmä**, jonka ideaa avaamme nyt ajatusleikin kautta.

Kuvitellaan, että olemme metsikköisessä laaksossa, jonka pohjalla on lähde, jonka luokse haluaisimme päästä. Maanpinnan korkeutta voi esimerkiksi ajatella kahden

muuttujan funktion kuvaajana, ja pyrimme etsimään tämän pinnan minimiarvoa eli laakson pohjaa. Puiden takia emme voi kuitenkaan päätellä, missä laakson pohja on tarkalleen, mutta voimme arvioida mistä suunnasta se voisi löytyä tutkimalla mihin suuntaan maaperä on kallistumassa lähellämme. Ilman muuta ylimääräistä tietoa on luonnollista lähteä etsimään laakson pohjaa kulkemalla siihen suuntaan, johon maa viettää. Gradienttimenetelmä perustuu ajatukseen, että otetaan pieniä askelia aina kohti sitä suuntaa, johon maanpinta laskee jyrkimmin. Tulkitaan tätä



Kuva 12. Gradienttimenetelmässä otetaan pieniä askelia funktion laskusuuntaan minimikohdan löytämiseksi.

nyt matemaattisesti, aloittaen yhden muuttujan tapauksesta.

Jos lähtöpisteemme on x_0 , niin funktion f derivaatasta $f'(x_0)$ nähdään, onko funktio f kasvava vai vähenevä tässä pisteessä. Derivaatan merkki siis kertoo, mihin suuntaan funktion kuvaaja johtaa alaspäin. Jos derivaatta on positiivinen, alaspäin pääsee kulkemalla x_0 :sta negatiiviseen suuntaan, ja päinvastoin jos $f'(x_0)$ on negatiivinen. Jos $f'(x_0) = 0$, ollaan tasanteella, ja toivon mukaan tällöin ollaan päädytty laakson pohjalle eli minimiarvoon. Gradienttimenetelmän tavoite onkin päästä tällaiselle tasanteelle, ottamatta vielä suurempaa kantaa onko tasanne aito minimi.

Mikäli ei olla vielä tasanteella, eli jos $f'(x_0) \neq 0$, halutaan lähteä pisteestä x_0 liikenteeseen ja ottaa siitä pieni askel uuteen pisteeseen x_1 . Luonnollinen tapa ottaa tämä askel on valita jokin pieni positiivinen luku $\varepsilon > 0$, ja lisätä pisteeseen x_0 luku $-\varepsilon f'(x_0)$, tuottaen kaavan

$$x_1 = x_0 - \varepsilon f'(x_0).$$

Tämä kaava takaa, että piste x_1 on valittu oikeasta suunnasta, sillä se valitaan aina siitä suunnasta mihin funktio f on laskeva. Lisäksi askeleen pituus on suhteessa funktion derivaatan suuruuteen, joka takaa luonnollisesti sen, että otamme

pienempiä askelia lähempänä maalia, eli derivaatan nollakohtaa.

Tätä prosessia voidaan jatkaa eteenpäin pisteestä x_1 , ja yleisesti ottaen gradienttimenetelmälle saadaan rekursiokaava

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

jota voidaan iteroida kunnes päästään riittävän lähelle derivaatan nollakohtaa. Askelpituuteen vaikuttavan parametrin ε voi valita ongelman ja omien tavoitteiden mukaisesti. Pienempi askelpituus voi vaatia enemmän laskemista, mutta päätyy ehkä varmemmin kohti derivaatan nollakohtaa.

Gradienttimenetelmän suppeneminen riippuu täysin tilanteesta. Jos menetelmä päätyy lähelle funktion f lokaalia minimiä, eli pistettä jonka ympärillä se saavuttaa vain suurempia arvoja, niin se tyypillisesti jää sinne jumiin eikä löydä mahdollista globaalia minimiarvoa. Tätä voi verrata laaksoesimerkissä kuoppaan kävelemisellä. Menetelmän suppenemisnopeus ja varmuus myös riippuu siitä, miten askelpituudet valitaan, ja tyypillisissä sovelluksissa askelpituus valitaan täysin ratkaistavan ongelman mukaisesti.

Usean muuttujan tapauksessa idea noudattaa samoja periaatteita – otetaan alkupisteestä $p_0 = (x_0, y_0)$ askel siihen suuntaan, mihin funktio on kaikista jyrkimmin laskeva. Tämän suunnan paljastaa gradientti $\nabla f(p_0) = \nabla f(x_0, y_0)$, sillä gradientti ilmaisee mihin suuntaan funktio kasvaa nopeiten (palaamme tämän väitteen todistukseen tehtävässä 38). Siispä sen vastavektori

$$-\nabla f(p_0) = \left(-\frac{\partial f}{\partial x}(p_0), -\frac{\partial f}{\partial y}(p_0) \right)$$

kertoo mihin suuntaan funktio laskee eniten. Voidaan jälleen muodostaa gradienttimenetelmälle rekursiokaava

$$p_{n+1} = p_n - \varepsilon \nabla f(p_n),$$

missä $\varepsilon > 0$ on edelleen vapaavalintainen parametri.

Gradienttimenetelmän aktiivisimmat sovellukset ovat useiden miljoonien muuttujien tapauksessa, sillä sitä sovelletaan erityisesti neuroverkkojen kouluttamisessa eli erilaisten tekoälyjen suunnittelussa ja luomisessa. Neuroverkkoa voi ajatella monimutkaisena laitteena, jonka ominaisuudet riippuvat lukemattomista eri muuttujista, joita halutaan säätää sopivan käyttäytymisen saavuttamiseksi. Kun ratkaistava ongelma on esimerkiksi kissan tunnistaminen valokuvasta, voidaan tekoälyä ensin pyytää luokittelemaan haluttu kuva, ja luokittelussa tapahtunutta virhettä voidaan ajatella tekoälyn toimintaa hallitsevien muuttujien funktiona. Tekoälyn kouluttamisessa on kyse tämän virheen minimoimisesta, ja mahdollista minimiarvoa voi yrittää lähestyä esimerkiksi gradienttimenetelmällä. Tekoälyn kouluttaminen on juurikin lukuisten pienten askelien ottamista, eli muuttujien säätämistä kohti pienintä virhearvoa.

Harjoitustehtäviä

Tehtävä 29. Selvitä, ovatko seuraavat funktiot konvekseja tai konkaaveja:

- a) $f(x) = x^2 - 4x + 1$
- b) $g(x) = x^3 + x^2$
- c) $h(x) = 1/x$ (kun $x > 0$)
- d) $k(x) = e^x$
- e) $l(x) = \ln(x)$ (kun $x > 0$)
- f) $m(x) = |x|^3$

Vihje s. 157 →

Toinen vihje s. 171 →

Tehtävä 30. Funktiolle $f(x)$ pätee $f(0) = 0$, $f(1) = 3$, ja $f(3) = 8$. Päättele, että f ei voi olla konvekssi funktio.

Vihje s. 157 →

Toinen vihje s. 171 →

Tehtävä 31. Todista konveksien funktioiden ensimmäisestä perusominaisuudesta (s. 50) toinen suunta: Näytä, että jokaisen konveksin derivoituvan funktion derivaatta on kasvava.

Vihje s. 157 →

Toinen vihje s. 171 →

Tehtävä 32. Todista neljän muuttujan aritmeettis-harmoninen epäyhtälö

$$\frac{4}{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \leq \frac{a + b + c + d}{4},$$

missä $a, b, c, d > 0$, soveltavalla Jensenin epäyhtälöä funktioon $f(x) = \frac{1}{x}$.

Vihje s. 158 →

Toinen vihje s. 172 →

Tehtävä 33. Sovella Jensenin epäyhtälöä funktiolla $f(x) = \ln(x)$, ja todista aritmeettis-geometrinen epäyhtälö

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n},$$

kun $x_1, \dots, x_n \geq 0$ ja $n \geq 1$.

Vihje s. 158 →

Toinen vihje s. 172 →

Tehtävä 34. Osoita, että seuraava **Youngin epäyhtälö** (William Henry Young, 1863–1942)

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

pätee kaikille positiivisille luvuille a, b ja p, q , missä $\frac{1}{p} + \frac{1}{q} = 1$.

Vihje s. 158 →

Toinen vihje s. 172 →

Tehtävä 35. Olkoot $x_1, \dots, x_n \geq 0$ ja $p > q > 0$. Osoita epäyhtälö

$$\left(\frac{x_1^p + x_2^p + \dots + x_n^p}{n} \right)^{\frac{1}{p}} \geq \left(\frac{x_1^q + x_2^q + \dots + x_n^q}{n} \right)^{\frac{1}{q}}$$

soveltamalla Jensenin epäyhtälöä sopivalle funktiolle.

Vihje s. 158 →

Toinen vihje s. 172 →

Tehtävä 36. Olkoon $0 < a < b$. Osoita Jensenin epäyhtälön integraaliversiota käyttäen, että

$$\frac{b-a}{\ln(b) - \ln(a)} \geq \sqrt{ab}.$$

Vihje s. 158 →

Toinen vihje s. 173 →

Tehtävä 37. Olkoon n positiivinen kokonaisluku. Osoita epäyhtälö

$$\sqrt{1^2 + 1} + \sqrt{2^2 + 1} + \dots + \sqrt{n^2 + 1} \geq \frac{n}{2} \sqrt{n^2 + 2n + 5}$$

soveltamalla Jensenin epäyhtälöä sopivalle funktiolle.

Vihje s. 158 →

Toinen vihje s. 173 →

Tehtävä 38. Olkoon $f(x, y)$ kahden muuttujan funktio, joka on derivoituva pisteessä (x_0, y_0) . Tällöin sille pätee määritelmän mukaan ensimmäisen asteen Taylorin kehitelmä (10) pisteessä (x_0, y_0) . Oletetaan, että gradientti $\nabla f(x_0, y_0)$ ei ole tässä pisteessä nollavektori. Todista, että gradientti osoittaa juuri siihen suuntaan, johon funktiota approksimoiva affiini kuvaus

$$L_{(x_0, y_0)}(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0)$$

kasvaa nopeiten.

Vihje s. 158 →

Toinen vihje s. 173 →

5 Numeerinen yhtälönratkaisu

Tutkitaan seuraavaa ongelmaa. Olkoon yhden muuttujan funktio f annettu, ja pyritään etsimään likiarvoista ratkaisua yhtälölle

$$f(x) = 0, \quad (13)$$

eli halutaan etsiä funktiolla f nollakohta. Tämä ongelma on varsin yleinen, sillä jokainen yhtälö, jossa esiintyy vain yhtä muuttujaa, voidaan palauttaa muotoon (13) siirtämällä kaikki lausekkeet yhdelle puolelle yhtälöä ja määrittelemällä funktio $f(x)$ edustamaan muodostunutta lauseketta. Esimerkiksi yhtälön $x^4 + 1 = x$ tapauksessa voidaan asettaa $f(x) = x^4 - x + 1$ tai vaikkapa $f(x) = \frac{x}{x^4 + 1} - 1$.

Esitämme seuraavaksi kaksi standardia numeerista menetelmää likiarvoisen ratkaisun löytämiseen.

5.1 Puolitusmenetelmä

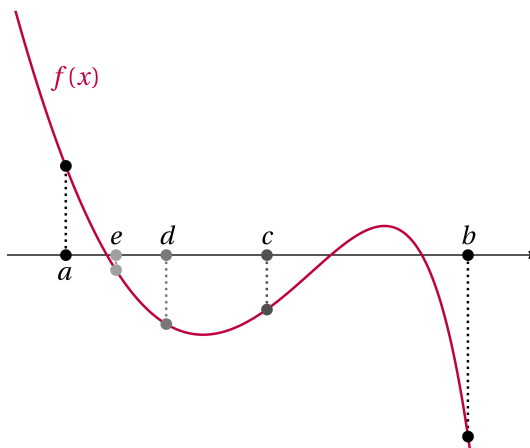
Tätä menetelmää varten oletamme, että annettu funktio f on jatkuva, ja että pysymme laskemaan funktion f arvoja. Puolitusmenetelmää varten tarvitsemme myös kaksi pistettä a ja b , joissa funktio f saa erimerkkiset arvot, eli joissa esimerkiksi $f(a)$ on positiivinen ja $f(b)$ on negatiivinen. Oletamme myös, että $a < b$. Voidaan hetkellisesti vaikkapa kuvitella, että tällaiset alkupisteet a ja b on löydetty syöttämällä funktioon f erilaisia arvoja, kunnes erimerkkiset kohdat on löydetty.

Kun alkupisteet a ja b on valittu, voidaan lähteä haarukoimaan funktiolla f nollakohtaa niiden välistä. Nollakohdan olemassaolo perustuu Luvussa 3 esittämäämme **Bolzanon lauseeseen**, jonka mukaan jatkuva funktio saa nollakohdan välillä, jonka päätepisteissä sillä on erimerkkiset arvot.

Kokeillaan ensin, löytyykö nollakohta välin keskipisteestä $c = (a + b)/2$. Mikäli c ei ole nollakohta, täytyy arvon $f(c)$ olla joko positiivinen tai negatiivinen. Riippuen kumpi tapaus toteutuu, voidaan prosessia jatkaa joko välillä $[a, c]$ tai $[c, b]$, sillä funktiomme saa jommassakummassa näistä kahdesta välistä erimerkkiset arvot välin päätepisteissä.

Edellä tapahtuneessa prosessissa olemme onnistuneet haarukoimaan nollakohdan pienemmälle välille kuin alkuperäinen väli $[a, b]$. Tätä prosessia voidaan toki toistaa, ja jokaisella kerralla nollakohdan sisältävän välin pituus puolittuu. Haarukoitavan välin pituus pienenee siis hyvin nopeasti, itse asiassa eksponentiaalisesti, sillä n iteraation jälkeen välin pituus on enää $(b - a)/2^n$.

Nollakohdan likiarvoksi voidaan käyttää esimerkiksi pienimmän löytyneen välin keskipistettä tai päätepisteitä – mitä pienempään väliin päädytään, sitä lähempänä nollakohtaa molemmat päätepisteet ovat.



Kuva 13. Puolitusmenetelmällä saadaan aikaan jono pisteitä (kuvassa $a, b, c, d, e, \dots$), joka suppenee kohti funktion nollakohtaa.

Puolitusmenetelmä, esimerkki. Lasketaan likiarvo luvulle $\sqrt{5}$ etsimällä funktion $f(x) = x^2 - 5$ positiiviselle nollakohdalle likiarvo puolitusmenetelmällä. Valitaan alkupisteiksi $a = 0$ ja $b = 3$, sillä funktio saa näissä pisteissä erimerkkiset arvot $f(0) = -5 < 0$ ja $f(3) = 4 > 0$.

Seuraava tarkasteltava piste on keskipiste $\frac{a+b}{2} = 1,5$, jossa funktiomme saa arvon $f(1,5) = -2,75 < 0$. Siis jatkamme prosessia välillä $[1,5; 3]$, jonka päätepisteissä funktio saa edelleen erimerkkiset arvot.

Välin $[1,5; 3]$ keskipiste on $2,25$, jossa funktio saa arvon $f(2,25) = 0,0625 > 0$. Tämä on jo melko lähellä nollaa, mutta jatketaan iterointia. Seuraava tutkittava väli on $[1,5; 2,25]$, jonka keskipisteessä $1,875$ funktiomme saa arvon $f(1,875) = -1,484 \dots < 0$. Prosessi jatkuu tästä seuraavan taulukon mukaisesti.

vasen päätepiste	funktion arvo	oikea päätepiste	funktion arvo
1,875	-1.484...	2,25	0,0625
2,0625	-0,746...	2,25	0,0625
2,15625	-0.350...	2,25	0,0625
2,22656255	-0,0424...	2,25	0,0625
2,2265625	-0,0424...	2,23828125	0,0099...

Tätä prosessia voidaan jatkaa, ja saadaan tarkempia ja tarkempia arvioita funktion todelliselle (positiiviselle) nollakohdalle $x = \sqrt{5} \approx 2,236068 \dots$

Puolitusmenetelmä on yksinkertainen ja tehokas menetelmä kun se pääsee käyntiin. Sopivien alkupisteiden a , b löytäminen saattaa kuitenkin olla vaikeaa tai jopa mahdotonta, jos funktio f saa pääsääntöisesti tietynmerkkisiä arvoja. Esimerkiksi funktio $f(x) = x^2$ ei saa negatiivisia arvoja lainkaan, joten sen nollakohdan löytäminen puolitusmenetelmällä ei yksinkertaisesti onnistu. Ja vaikka puolitusmenetelmä suppenee eksponentiaalisella vauhdilla, on seuraavaksi esittelemämme Newtonin menetelmä tyypillisesti paljon nopeampi.

5.2 Newtonin menetelmä



Isaac Newton (1643–1727)

Oletetaan nyt lisäksi, että funktio f on jatkuvasti derivoituva, ja että pystymme laskemaan sekä f :n että sen derivaatan arvoja halutessamme. Newtonin menetelmää varten tarvitsemme myös yhden aloituspisteen x_0 , joka voi olla mikä tahansa valitsemamme piste.

Tutkitaan nyt tilannetta pisteessä x_0 . Aloitetaan laskemalla arvot $f(x_0)$ ja $f'(x_0)$. Arvo $f(x_0)$ kertoo funktion f suuruuden pisteessä x_0 , ja voidaan olettaa, että x_0 ei ole nollakohta, sillä muuten menetelmä on jo onnistunut. Derivaatan arvo $f'(x_0)$ taas kertoo funktion kasvunopeuden pisteessä x_0 , eli mihin suuntaan funktio on menossa pisteessä x_0 . Tämä antaa vähän ideaa kummasta suunnasta kannattaa lähteä etsimään nollakohtaa: Mikäli esimerkiksi arvo $f(x_0)$ on positiivinen ja $f'(x_0)$ on negatiivinen, kannattaa valita seuraava arvauspiste x_1 arvon x_0 oikealta puolelta, eli niin, että $x_1 > x_0$, sillä tähän suuntaan funktion f arvot ovat todennäköisemmin lähempänä nollaa.

Tilanne voi toki olla todellisuudessa monimutkaisempi, mutta ei murehdi siitä vielä. Sen sijaan pohditaan, kuinka pitkältä seuraava arvaus x_1 kannattaisi valita. Tähän Newtonin menetelmä tarjoaa idean: muodostetaan funktiolle f pisteessä x_0 tangenttisuoran yhtälö, joka saadaan kaavasta

$$y = f'(x_0)(x - x_0) + f(x_0),$$

ja asetetaan seuraava arvaus x_1 tämän suoran nollakohtaksi. Asettamalla $y = 0$ ja ratkaisemalla suoran yhtälö x :n suhteen saadaan

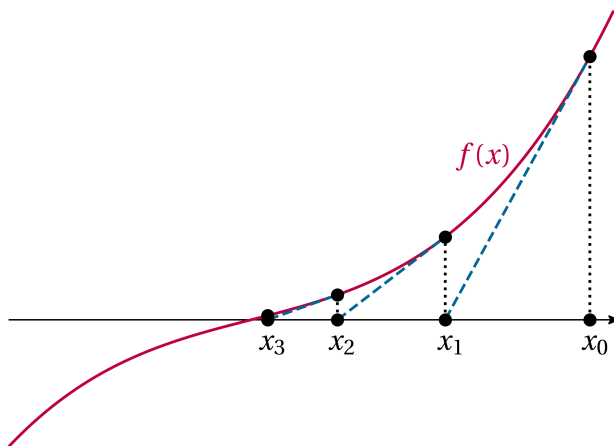
$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Toki tätä menettelyä varten vaaditaan myös, että derivaatan arvo $f'(x_0)$ ei ole nolla, sillä muuten tangenttisuora on vaakasuora ja sillä ei ole nollakohtaa. Tällaisessa tapauksessa täytyy keksiä uusi alkuarvaus x_0 .

Prosessi voidaan nyt aloittaa uudelleen alusta pisteestä x_1 lähtien, ja tuottaa uusi piste x_2 , joka taas on nollakohta funktion f tangentille pisteessä x_1 . Toistamalla tätä toimenpidettä saadaan arvauspisteille rekursiokaava

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad \text{kun } n = 0, 1, 2, \dots,$$

joka tuottaa jonon pisteitä $x_0, x_1, x_2, \dots$. Toive on, että tämä pistejono lähestyy jotakin funktion f nollakohtaa. Tutkitaan nyt tapahtuuko näin käytännössä.



Kuva 14. Newtonin menetelmän kolme ensimmäistä askelta visualisoituna.

Newtonin menetelmä, esimerkki 1. Aloitetaan esimerkillä $f(x) = x^2$, johon puolitusmenetelmä ei tepsinyt. Koska $f'(x) = 2x$, saadaan rekursiokaavaksi tässä tapauksessa

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

Rekursiokaavan perusteella nähdään, että Newtonin menetelmässä seuraava jäsen on aina puolet edellisestä, eli alkuarvosta riippumatta menetelmä suppenee aina kohti nollakohtaa $x = 0$.

Newtonin menetelmä, esimerkki 2. Etsitään Newtonin menetelmällä ratkaisua yhtälölle $\cos(x) = x$. Approksimaatio tulee tarpeeseen, sillä ratkaisua ei voida esittää suljetussa muodossa. Valitaan funktioksi $f(x) = \cos(x) - x$. Funktion derivaatta on $f'(x) = -\sin(x) - 1$, joten iteraatiokaavasta tulee

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n + \frac{\cos(x_n) - x_n}{\sin(x_n) + 1}.$$

Aloitetaan alkuarvauksella $x_0 = 1$. Saadaan seuraavanlainen jono pisteitä

n	x_n	$f(x_n)$
0	1	-0.459...
1	0,750363...	-0,0189...
2	0,739112890...	$-4,645 \dots \cdot 10^{-5}$
3	0,739085133385...	$-2,847 \dots \cdot 10^{-10}$
4	0,739085133215...	≈ 0
5	0,739085133215...	≈ 0

Neljännän iteraation jälkeen funktion arvot ovat jo niin lähellä nollaa, että laskimen 15 desimaalin tarkkuus ei enää riitä ja muuttujan x arvo ei enää muutu. Newtonin menetelmä suppenee siis hyvin nopeasti tässä tapauksessa.

Yhtälön ratkaisu on siis esimerkiksi kuuden desimaalin tarkkuudella $x \approx 0,739085$. Jos halutaan todistaa, että tämä todella on oikea likiarvo, voidaan käyttää Bolzanon lausetta hieman tätä likiarvoa suurempaan ja pienempään lukuun: koska

$$\begin{aligned} f(0,7390854) &= -4,464 \dots \cdot 10^{-7} < 0 \text{ ja} \\ f(0,7390846) &= 8,923 \dots \cdot 10^{-7} > 0, \end{aligned}$$

Bolzanon lause takaa, että välillä $[0,7390846; 0,7390854]$ on funktion f nollakohta. Koska *kaikki* luvut tältä väliltä pyöristyvät kuuden desimaalin tarkkuudella arvoon 0,739085, tämä on myös funktion nollakohdan oikea likiarvo.

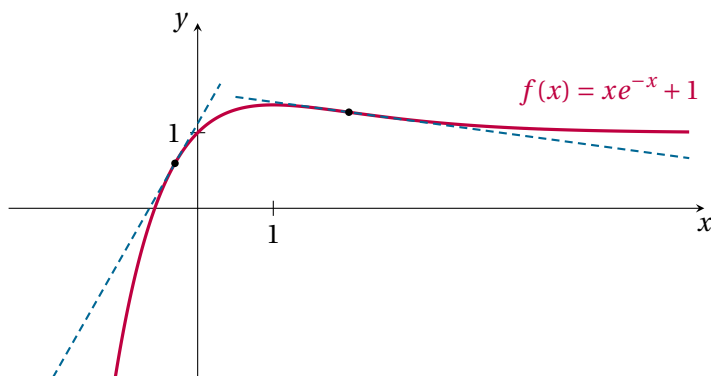
Newtonin menetelmän suppeneminen

Newtonin menetelmän tuottaman pistejonon suppeneminen kohti funktion nollakohtaa ei ole mitenkään ilmiselvä kysymys. Menetelmä toimii käytännössä melko

usein, mutta sillä on myös omat oikkunsa, joihin liittyy mielenkiintoisia matemaattisia kysymyksiä.

Tutkitaan ensin tilanteita, joissa Newtonin menetelmä toimii varmasti. Oletetaan, että funktiolla f on nollakohta a , funktio f on derivoituva pisteen a läheisyydessä, sen derivaattafunktio f' on täällä jatkuva, ja derivaatan arvo nollakohdassa on nolasta poikkeava, $f'(a) \neq 0$. Tässä tapauksessa Newtonin menetelmä suppenee varmasti, kunhan alkuarvo x_0 on valittu riittävän läheltä nollakohtaa a . Tämän tuloksen todistus löytyy liitteestä E (s. 148).

Riippuu funktiosta f , millaiset alkuarvot ovat riittävän lähellä. Esimerkiksi funktion $f(x) = xe^{-x} + 1$ tapauksessa iteraatio suppenee ainoaan nollakohtaan $x \approx -0,567$ aina, kun $x_0 < 1$, mutta kun $x_0 > 1$, iteraatio hajaantuu kohti ääretöntä.



Neliöllinen suppeneminen. Otamme seuraavaksi kantaa Newtonin menetelmän suppenemisvauhtiin. Osoittautuu, että jos edellisiin oletuksiin lisätään vielä pari lisäehtoa, voidaan Newtonin menetelmälle osoittaa hyvin nopea ns. *neliöllinen* suppenemisvauhti.

Lisätään siis oletuksiin vielä, että $f'(a) \neq 0$, ja oletetaan myös, että funktion f toinen derivaatta f'' on olemassa pisteen a lähetyvillä ja sille pätee siellä yläarvio $|f''(x)| \leq M$ jollakin vakiolla M . Tällöin voidaan todistaa, että

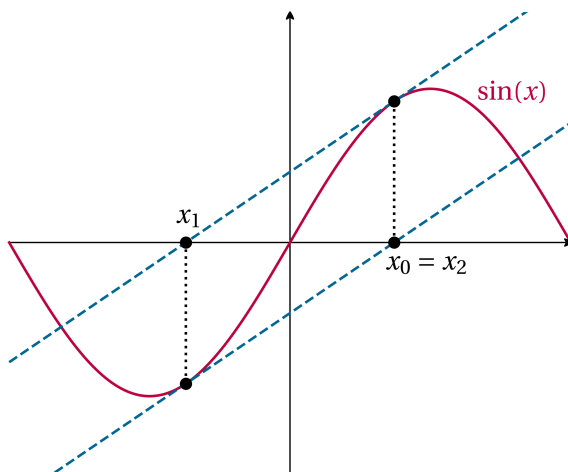
$$|x_{n+1} - a| \leq \frac{3M}{|f'(a)|} |x_n - a|^2, \quad \text{kun } n \geq 0. \quad (14)$$

Tämä arvio on melko voimakas, sillä se kertoo, että Newtonin menetelmässä seuraavan iteraatiopisteen absoluuttinen virhe nollakohtaan a on aina verrannollinen edellisen virheen neliöön. Jos virhe on jo pieni, sen neliö on vielä paljon pienempi. Tällainen suppeneminen on huomattavasti eksponentiaalistakin nopeampaa, ja peittoaa siis helposti esimerkiksi puolitusmenetelmän. Myös arvion (14) todistus löytyy liitteestä E.

Newtonin menetelmän suppenemattomuus

Newtonin menetelmä voi myös epäonnistua, esimerkiksi suppenematta lainkaan tai jopa suppenemalla väärään lukuun. Tutkitaan nyt mahdollisia tilanteita ja syitä, miksi näin voi käydä.

Tilanne 1: Huono alkuarvaus. Jos alkuarvausta x_0 ei ole valittu riittävän läheltä funktion nollakohtaa, voi yksinkertaisesti käydä niin, että jono x_n käyttäytyy huonosti ja rupeaa vaikkapa toistamaan itseään. Näin voi käydä esimerkiksi sinifunktiolle $f(x) = \sin(x)$, jolle Newtonin menetelmä voi tietyllä alkuarvauksella jäädä yksinkertaisesti palloilemaan kahden pisteen välille (tehtävä 44).



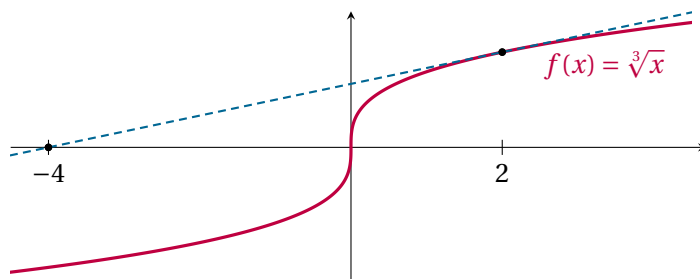
Kuva 15. Sopivalla alkupisteen valinnalla Newtonin menetelmä sahaa kahden pisteen välillä.

Kuvassa esiintyneen alkupisteen määrittämistä tutkitaan tehtävissä. Tämä on suhteellisen yleinen ilmiö Newtonin menetelmälle, mutta käytännössä ongelma yleensä ratkeaa vaihtamalla alkuarvausta, kunnes suppeneva jono löytyy.

Tilanne 2: Funktio ei derivoituva. Jos funktio f ei ole kaikkialla derivoituva, voidaan joutua vaikeuksiin. Esimerkiksi funktiolle $f(x) = \sqrt[3]{x}$ Newtonin menetelmä ei tepsii ollenkaan, sillä rekursiokaavasta saadaan

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^{\frac{1}{3}}}{\frac{1}{3}x_n^{-\frac{2}{3}}} = x_n - 3x_n = -2x_n.$$

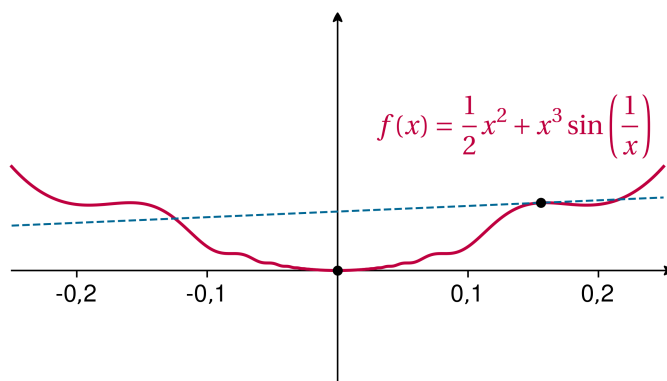
Tästä nähdään, että lukujen x_n muodostama jono ei koskaan suppene kohti f :n ainoaa nollakohtaa $x = 0$, sillä jonon jokainen jäsen on kaksi kertaa kauempana nolasta kuin edellinen. Ongelma johtuu siitä, että funktio f ei ole derivoituva nollakohdassaan.



Tilanne 3: Liian litteä kuvaaja. Selvitetään vielä, miksi aiemmin käsitellyissä suppenemistarkasteluissa oli tarpeellista olettaa, että $f'(a) \neq 0$ funktion nollakohdassa a . Oheisessa kuvassa on hahmoteltu funktio

$$f(x) = \begin{cases} \frac{1}{2}x^2 + x^3 \sin\left(\frac{1}{x}\right), & \text{kun } x \neq 0 \\ 0, & \text{kun } x = 0 \end{cases},$$

jolla on nollakohta pisteessä $x = 0$.



Tämä funktio on lisäksi jatkuvasti derivoituva, ja sen derivaatalla on nollakohta pisteessä $x = 0$. Osoittautuu myös, että tämän nollakohdan lisäksi derivaatalla on useita muitakin nollakohtia nollan läheisyydessä. Itse asiassa mielivaltaisen läheltä nollaa löytyy aina uusia derivaatan nollakohtia.

Jos Newtonin menetelmän alkuarvaus x_0 valitaan liian läheltä tällaista nollakohtaa, niin kuvaajalle piirretty tangenttisuora on lähes vaakasuora, ja seuraava iteraatiopiste x_1 päättyy hyvin kauas nollasta.

Tämä esimerkki osoittaa, että vaikka alkuarvaus x_0 valittaisiin kuinka läheltä nollakohtaa $x = a$, on silti mahdollista, että Newtonin menetelmä karkaa kauemmaksi, jos $f'(a) = 0$. Tässä ongelman aiheuttaja on se, että funktion derivaatta f' sai arvoja liian läheltä nollaa alkuperäisen funktion f nollakohdan $x = 0$ ympäristössä.

Oleellisesti samaa ideaa käyttäen voidaan tuottaa myös villimpiä esimerkkejä, sillä muokkaamalla funktion kuvaajaa iteraatiopisteen läheisyydessä voidaan saavut-

taa mikä tahansa haluttu käyttäytyminen ainakin yhden tietyn alkuarvauksen tuottamalle jonolle. On esimerkiksi mahdollista löytää sellainen funktio ja alkuarvaus, että Newtonin menetelmä suppenee kohti pistettä, joka ei ole funktion nollakohta (tehtävä 45).

5.3 Kiintopisteiteraatio

Newtonin menetelmää voi tulkita myös erikoistapauksena yleisemmästä yhtälönratkaisumenetelmästä nimeltä **kiintopisteiteraatio**. Kiintopisteiteraation tavoite on ratkaista yhtälö

$$g(x) = x, \quad (15)$$

missä g on jokin annettu jatkuva funktio. Mikäli x toteuttaa tämän yhtälön, sitä sanotaan funktion g **kiintopisteeksi**, eli pisteeksi joka ei muutu funktiota g sovellettaessa. Toki yleisen yhden muuttujan yhtälön muotoa $f(x) = 0$ voi aina palauttaa myös yhtälön (15) muotoon asettamalla $g(x) = f(x) + x$.

Kiintopisteiteraation tarjoama ratkaisumenetelmä yhtälölle (15) perustuu funktion g iteroimiseen. Tarkemmin sanottuna, alkuarvauksesta x_0 muodostetaan jono pisteitä rekursiokaavalla

$$x_{n+1} = g(x_n), \quad \text{kun } n = 0, 1, 2, \dots, \quad (16)$$

eli funktiota g sovelletaan toistuvasti. Siispä $x_n = g(g(\dots g(x_0)\dots))$, missä funktiota g on sovellettu n kertaa. Toive on, että tämä jono suppenee, sillä jos näin muodostetulla jonolla (x_n) on raja-arvo x , niin pisteen x täytyy olla yhtälön (15) ratkaisu. Tämä on helppo perustella ottamalla raja-arvo puolittain yhtälöstä (16), sillä tällöin saadaan

$$x = \lim_{n \rightarrow \infty} x_{n+1} = \lim_{n \rightarrow \infty} g(x_n) = g\left(\lim_{n \rightarrow \infty} x_n\right) = g(x),$$

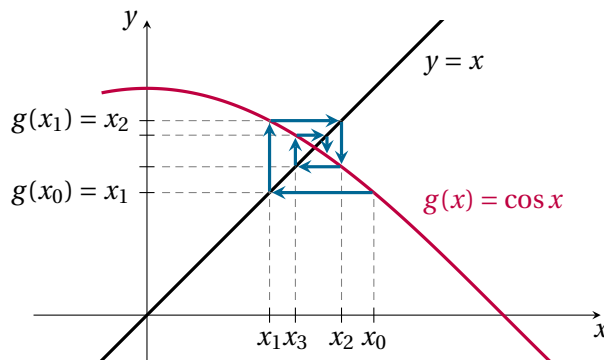
missä hyödynsimme funktion g jatkuvuutta toiseksi viimeiseen yhtäsuuruuteen.

Kiintopisteiteraatio, esimerkki. Tutkimme aiemmin esimerkin kautta yhtälön $\cos(x) = x$ ratkaisemista Newtonin menetelmällä. Ratkaistaan sama yhtälö nyt kiintopisteiteraation avulla. Iteraatiofunktio on sopivasti $g(x) = \cos(x)$.

Aloittamalla iteraatio pisteestä $x_0 = 1$ saadaan seuraava taulukko. Virhettä on verrattu todelliseen kiintopisteeseen $x \approx 0,739085$, jonka selvitimme aiemmin Newtonin menetelmällä.

jäsen	arvo	virhe	suhteellinen virhe
x_0	1	0,2609...	35 %
$x_1 = g(x_0)$	0,5403...	-0,1987...	27 %
$x_2 = g(x_1)$	0,8575...	0,1184...	16%
x_3	0,6542...	-0,0847...	11 %
x_4	0,7934...	0,0543...	7,4 %
x_5	0,7013...	-0,0377...	5,1 %
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{50}	0,7390...	$7,06 \dots \cdot 10^{-10}$	$9,6 \cdot 10^{-12} \%$

Kiintopisteiteraation voi visualisoida miellyttävästi:



Huomataan, että iteraatiopisteet lähenevät todellista kiintopistettä vakaasti, mutta eivät erityisen nopeasti. Suhteellinen virhe pienenee suunnilleen geometrisesti, ja lähellä kiintopistettä tämä suhde lähenee lukua $|f'(0,739085\dots)| \approx 0,67$. Seuraava teoreettinen tulos kertoo, miksi niin kuuluukin olla.

Kiintopisteiteraation suppenemisehto 1. Jos funktio g on derivoituva, ja löytyy vakio $k < 1$, jolla sen derivaatalle pätee $|g'(x)| \leq k$ kaikilla $x \in \mathbb{R}$, niin funktiolla g on kiintopiste $x = g(x)$, ja kiintopisteiteraatio suppenee kohti tätä kiintopistettä.

Todistetaan lause seuraavalla tavalla. Väitämme ensin, että ehdon $|g'(x)| < k$ toteuttava funktio toteuttaa itse asiassa yleisemmän ehdon

$$|g(x) - g(y)| \leq k|x - y| \quad \text{kaikilla } x, y. \quad (17)$$

Ehdon (17) toteuttavia funktiota (missä $k < 1$) sanotaan **kontraktoiksi**, eli ne ovat funktioita joissa kuvapisteen etäisyys on aina pienempi kuin lähtöpisteiden etäisyys.

Oletuksesta $|g'(x)| \leq k$ voidaan todistaa epäyhtälö (17) käyttämällä väliarvolausetta. Lause sanoo, että pisteiden x ja y välissä on olemassa piste ξ , jolle $g(x) - g(y) = g'(\xi)(x - y)$. Tästä saadaan suoraan, että

$$|g(x) - g(y)| = |g'(\xi)| |x - y| \leq k|x - y|.$$

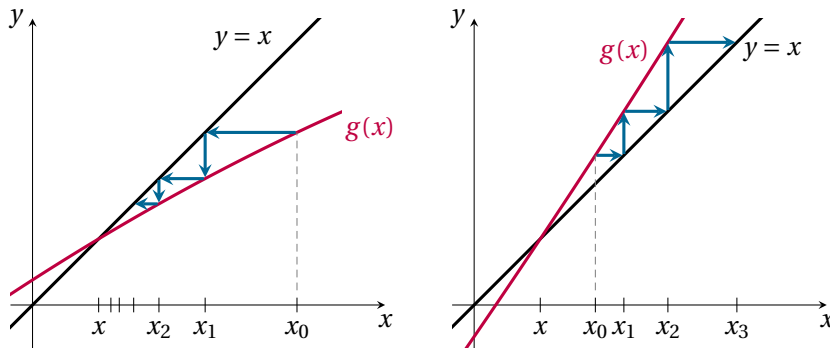
Palataan nyt jonon (x_n) suppenemisen määrittämiseen. Epäyhtälöä (17) soveltamalla havaitaan, että

$$|x_{n+1} - x_n| = |g(x_n) - g(x_{n-1})| \leq k|x_n - x_{n-1}|.$$

Siispä perättäisten jonon pisteiden etäisyys on enintään k kertaa edellisten pisteiden etäisyys. Koska $k < 1$, tämä etäisyys suppenee kohti nollaa eksponentiaalisen nopeasti. Tästä on mahdollista päätellä myös, että jono (x_n) suppenee eksponentiaalisella vauhdilla kohti jotakin raja-arvoa x . Kuten aiemmin havaitsimme, raja-arvopisteen x täytyy olla yhtälön (15) ratkaisu. Siispä menetelmä on tuottanut ratkaisun. $\square$

Huom! On myös mahdollista, että funktiolla on kiintopiste, vaikka sen derivaatta ei toteuttaisi lauseen ehtoa $|g'(x)| \leq k < 1$. Esimerkiksi funktiolla $g(x) = 2x$ on kiintopiste $x = 0$, mutta kiintopisteiteraation suora soveltaminen ei löydä sitä.

Yleisesti ottaen funktion derivaatan arvo $g'(x)$ kiintopisteessä x antaa informaatiota siitä, mitä funktiota g iteroitaessa tapahtuu pisteen x läheisyydessä. Mikäli $g'(x) < 1$, on kiintopiste x funktion **vakaa kiintopiste**. Tämä tarkoittaa erityisesti sitä, että läheltä pistettä x valitut alkupisteet x_0 suppenevat edelleen kohti kiintopistettä x funktiota iteroitaessa. Siis pieni muutos pisteeseen x ei muuta iteraatiojonon $x_0, g(x_0), g(g(x_0)), \dots$ käyttäytymistä.



Sen sijaan jos $g'(x) > 1$, kiintopiste x on **epävakaa kiintopiste**. Jos tässä tapauksessa alkuarvaus x_0 on valittu läheltä pistettä x , niin jono $x_0, g(x_0), g(g(x_0)), \dots$ karkaa aina pois päin kiintopisteestä x , ainakin lokaalisti. Teoriassa jono voi edelleen palata lähelle pistettä x riippuen funktion käyttäytymisestä kauempana, mutta epävakaa kiintopisteen tapauksessa funktio työntää aina pisteen lähellä olevia arvoja kauemmas kiintopisteestä. Pieni muutos kiintopisteeseen x aiheuttaa siis huomattavan eron iteraatiojonon käyttäytymisessä.

Kiintopisteiteraation käyttäytymistä ei kuitenkaan voi aina päätellä pelkästään derivaatan avulla. Jos kiintopisteessä x pätee $g'(x) = 1$, tai jos funktio g ei ole derivoituva pisteessä x , niin yleisesti ei voida sanoa enempää kiintopisteen laadusta.

Mikäli kiintopisteen x läheisistä pisteistä iterointi edelleen suppenee pisteeseen x tai karkaa siitä kauemmas, voidaan piste luokitella taas vakaaksi tai epävakaaksi kiintopisteeksi. On myös tapauksia, jotka eivät sovi kumpaankaan näistä luokista. Seuraava tulos antaa riittävän ehdon sille, että kiintopisteiteraatio suppenee kohti vakaata kiintopistettä alkuarvolla x_0 .

Kiintopisteiteraation suppenemisehto 2

Olkoon funktiolla g kiintopiste $a = g(a)$. Kiintopisteiteraatio suppenee lukuun a alkuarvauksella x_0 , mikäli pätee

$$|g'(x)| \leq k < 1,$$

kun x on lähempänä tai yhtä kaukana pisteestä a kuin piste x_0 .

Todistus. Tutkitaan kiintopisteiteraatiota $x_{n+1} = g(x_n)$, missä $n = 0, 1, 2, \dots$. Ensimmäisen suppenemisehdon todistuksen tapaan voidaan arvioida väliarvolauseella

$$|x_1 - a| = |g(x_0) - g(a)| = |g'(\xi)| |x_0 - a| \leq k |x_0 - a|.$$

Piste x_1 on siis lähempänä kiintopistettä a kuin alkuarvo x_0 , eli iteraatio pysyy alueella, jossa g on kontraktio. Vastaavasti arvioiden saadaan

$$|x_n - a| \leq k |x_{n-1} - a| \leq \dots \leq k^n |x_0 - a|,$$

joten iteraatio suppenee kiintopisteeseen, sillä $k^n \xrightarrow{n \rightarrow \infty} 0$. □

Newtonin menetelmä kiintopisteiteraationa. Newtonin menetelmää voi myös tulkita kiintopisteiteraationa funktiolle $g(x) = x - \frac{f(x)}{f'(x)}$, koska menetelmä perustuu tämän funktion iteroimiseen. Luonnollisesti funktion g kiintopisteet ovat funktion f nollakohtia, sillä

$$g(x) = x \Rightarrow 0 = \frac{f(x)}{f'(x)} \Rightarrow 0 = f(x).$$

Käänteinen päätelmä voidaan tehdä myös, kun $f'(x) \neq 0$. Newtonin menetelmässä valittu lauseke funktiolle g on myös järkevä valinta nollakohdan löytämiseen, sillä jos funktio f on kahdesti derivoituva, niin voidaan laskea

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

Mikäli $f(a) = 0$ ja $f'(a) \neq 0$, huomataan, että $g'(a) = 0$. Tästä voidaan päätellä, että jokainen yhtälön $f(x) = 0$ ratkaisu, joka ei ole derivaatan nollakohta, on itse asiassa

funktion g vakaa kiintopiste. Tämä perustelee jälleen Newtonin menetelmän toimivuutta nollakohdan a läheisyydessä, kun $f'(a) \neq 0$.

Kiintopisteiteraation suppenemisehtoa $|g'(x)| \leq k < 1$ voidaan soveltaa suoraan edellä laskettuun derivaattaan, jolloin saadaan Newtonin menetelmälle ehto, joka takaa suppenemisen.

Newtonin menetelmän suppenemisehto

Newtonin menetelmä suppenee, jos funktiolla f on nollakohta, ja

$$\left| \frac{f(x) \cdot f''(x)}{(f'(x))^2} \right| \leq k < 1,$$

kun x on lähempänä tai yhtä kaukana nollakohtaa kuin alkuarvaus x_0 .

Tämä on erittäin käyttökelpoinen tulos, kun halutaan etukäteen varmistaa, että Newtonin menetelmä todella suppenee.

Havainnollistamme seuraavaksi Newtonin menetelmän kaoottisuutta geometrisesti, mutta tätä varten meidän täytyy yleistää menetelmä uuteen kontekstiin.

5.4 Newtonin dynamiikkaa kompleksitasossa

Newtonin menetelmää on mahdollista soveltaa myös kompleksitason funktioille, eli funktioille $f: \mathbb{C} \rightarrow \mathbb{C}$, jotka ottavat syötteeksi ja tulostavat arvoikseen kompleksilukuja. Tässä tapauksessa menetelmän geometrinen tulkinta kuvaajan tangenttisuoran seuraamisena ei enää ole järkevä, mutta monille kompleksitason funktioille on edelleen mahdollista määritellä kompleksinen derivaatafunktiio f' . Tällöin funktion

$$g(z) = z - \frac{f(z)}{f'(z)}$$

iteroimista jostain alkuarvauspisteestä z_0 lähtien voidaan edelleen ajatella Newtonin menetelmän yleistykseenä kompleksitasoon. Kompleksisen derivaatan omaavilla funktioilla on erityinen rooli matematiikassa, ja niitä sanotaan yleensä **kompleksiseksi analyyttiseksi** funktioiksi.

Emme mene tässä yksityiskohtiin kompleksisen derivaatan määrittelystä, sillä keskitymme ainoastaan kompleksisten polynomifunktioiden tutkimiseen. Polynomit voidaan määritellä kompleksitasossa tavalliseen tapaan, sillä kompleksilukuja voi laskea yhteen ja kertoa keskenään. Yleinen kompleksinen polynomifunktio astetta n on muotoa

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0,$$

missä luvut $a_0, \dots, a_n$ ovat kompleksilukuja, ja $a_n \neq 0$. Myös reaalikertoimisia polynomeja kuten $p(z) = z^3 + z^2 + z + 1$ on usein hyödyllistä tulkita kompleksitason

polynomeina. Kompleksisen polynomin derivaatta voidaan laskea aivan samalla tavalla kuin reaalisessa tapauksessa, esimerkiksi $p'(z) = 3z^2 + 2z + 1$.

Keskitymme nyt kompleksisen polynomin nollakohtien eli juurien etsimiseen. Kuten reaalisessa tapauksessa, polynomit voidaan edelleen jakaa tekijöihin juuriensa perusteella. Esimerkiksi polynomillamme $p(z) = z^3 + z^2 + z + 1$ on kolme juurta, -1 ja $\pm i$, ja se voidaan jakaa tekijöihin muotoon

$$p(z) = (z + 1)(z - i)(z + i).$$

Polynomien tekijöihinjako osoittaa, että polynomilla voi olla enintään asteensa verran juuria, moninkertaiset juuret mukaan luettuna. Osoittautuu kuitenkin, että jokaisella polynomilla on *tasan* asteensa verran juuria kompleksitasossa, edelleen moninkertaiset juuret mukaan luettuna. Tämä tulos tunnetaan **algebran peruslauseena**, joka on esitelty Lisäsivujen osassa 5. Jossain mielessä tulos kertoo, että kompleksitaso on luonnollisempi ympäristö polynomeille kuin reaaliakseli. Esimerkiksi polynomilla $q(x) = x^2 + 2$ ei ole reaalisia juuria, sillä ne piileskelevät kompleksitasossa pisteissä $\pm\sqrt{2}i$.

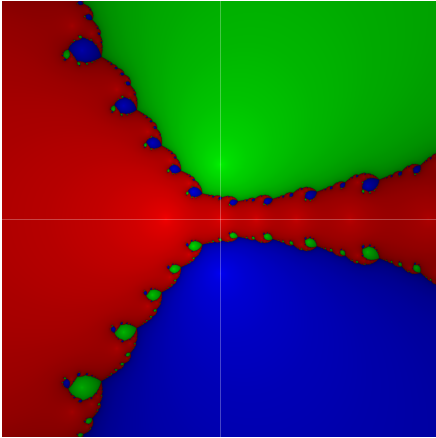
Tämä tulos tekee polynomin juurien etsimisestä kompleksitasossa myös erittäin mielenkiintoisen kysymyksen, sillä se takaa, että jokaisella polynomilla on ainakin yksi nollakohta (paitsi toki nollasta poikkeavilla vakiofunktioilla). On mahdollista myös näyttää, että polynomifunktioon sovellettuna Newtonin menetelmä tulee aina suppenemaan nollakohtien läheisyydessä, eli jokaisen polynomin f juuret ovat vakaita kiintopisteitä funktiolle $z - \frac{f(z)}{f'(z)}$. Vaikuttaa siis siltä, että Newtonin menetelmä on varsin luonnollinen vaihtoehto polynomien juurten etsimiseen.

Newtonin fraktaalit

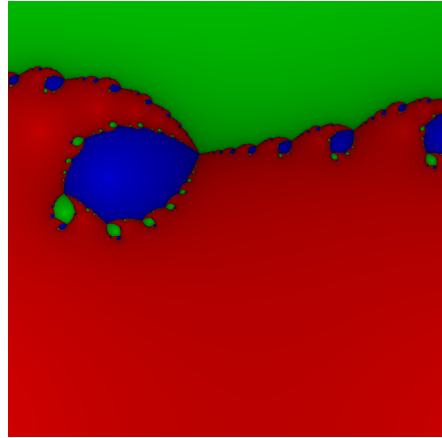
Mandelbrotin joukkoon (Pitkän matematiikan lisäsivut 5) tutustunut lukija saat-taa arvata, että kompleksitasossa funktioiden iterointi voi tuottaa monimutkaisia ja mielenkiintoisia lopputuloksia. Myös Newtonin menetelmän tutkiminen joh-taa kaoottisiin ja kauniisiin fraktaalirakenteisiin, joiden syntymistä esittelemme seuraavaksi.

Aloitetaan polynomifunktiosta $p(z) = z^3 + z^2 + z + 1$. Kuten aiemmin totesimme, polynomilla p on tasan kolme juurta, -1 ja $\pm i$. Osoittautuu, että tälle kyseiselle kolmannen asteen polynomille Newtonin menetelmä tepsii hyvin - lähes jokainen alkuarvauspiste z_0 suppenee johonkin näistä kolmesta juuresta. Tätä hyödyntäen lähdemme värittämään kompleksitason pisteet yhdellä kolmesta väristä (valitaan punainen, sininen, vihreä) sen perusteella, mihin kolmesta juuresta Newtonin menetelmä suppenee. Tätä on havainnollistettu Kuvissa 16 ja 17.

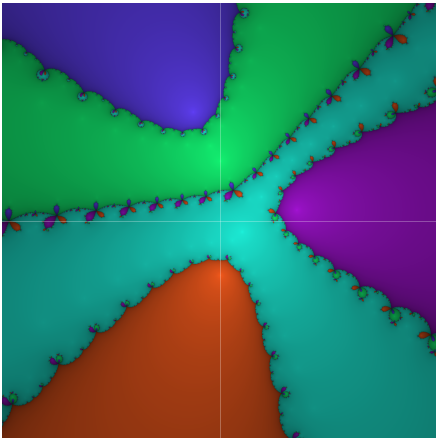
Näin syntyviä fraktaalikuvia kutsutaan **Newtonin fraktaaleiksi**, ja ne ovat kaunis tapa havainnollistaa menetelmän toimintaa. Lisää esimerkkejä on esitelty Kuvissa 18 ja 19.



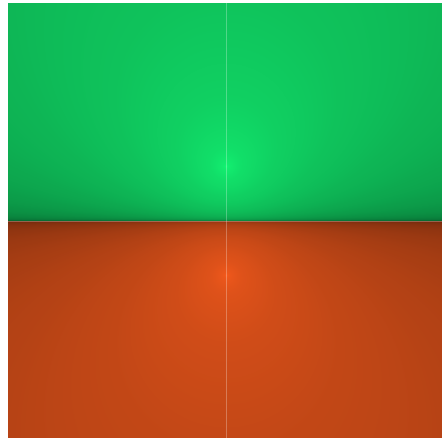
Kuva 16. Newtonin fraktaali polynomille $z^3 + z^2 + z + 1$. Väri kertoo, mihin juureen menetelmä suppenee. Tummemmat pisteet suppenevat hitaammin.



Kuva 17. Sama kuva suurennettuna lähelle alueiden reunaa. Fraktaalirakenne jatkuu loputtomiin.



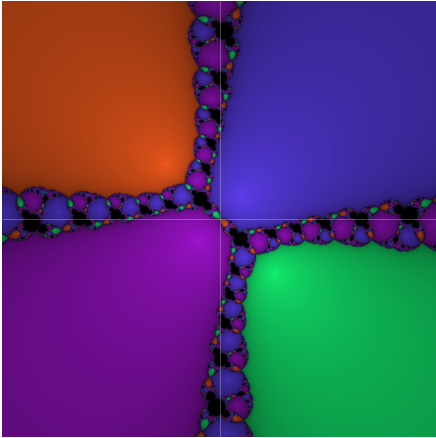
Kuva 18. Newtonin fraktaali erälle viidennen asteen polynomille. Jokaisen reunapisteen läheltä voidaan löytää jokaiseen eri juureen suppenevia pisteitä.



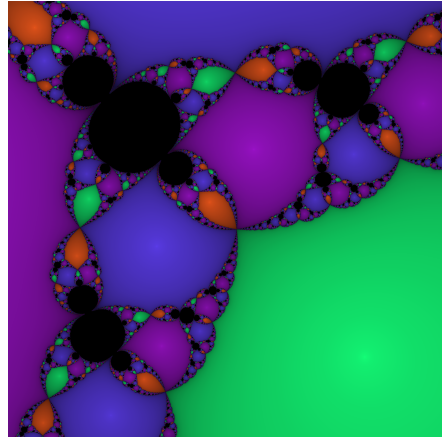
Kuva 19. Toisen asteen polynomille fraktaalirakennetta ei synny. Kuvassa polynomi $z^2 + 1$.

Tutkitaan vielä pisteitä, joissa Newtonin menetelmä ei suppene. Edellä käsitellyissä tapauksissa kävi niin, että tällaisten suppenemattomien pisteiden joukot olivat häilyvän ohuita. On kuitenkin mahdollista, että Newtonin menetelmä ei suppene laajemmassa alueessa, jopa yksinkertaisille polynomeille. Kuvissa 20 ja 21 on esitetty tällainen tapaus, missä suppenemattomat pisteet on väritetty mustaksi.

Myös mustissa alueissa näkyy fraktaalirakennetta, ja niiden muodossa on sa-



Kuva 20. Newtonin fraktaali eräälle neljännisen asteen polynomille. Mustan alueen pisteet eivät suppene.



Kuva 21. Sama kuva suurennettuna mustan alueen lähelle.

moja piirteitä kuin Mandelbrotin joukossa. Näiden fraktaalien rakenteeseen ja yhteyksiin Mandelbrotin joukon kanssa liittyy paljon syvällistä matematiikkaa, jonka käsittelyyn tarvittavia esitietoja emme tässä ehdi käymään läpi. Mainittakoon kuitenkin, että näitä aiheita käsittelevä matematiikan osa-alue on nimeltään **kompleksianalyysi** ja erityisesti sen osa-alue **kompleksidynamiikka**.

Harjoitustehtäviä

Osa tehtävistä saattaa vaatia laskinta, tai esimerkiksi lyhyttä koodinpätkää tietokoneella Newtonin menetelmän toteuttamiseen.

Tehtävä 39. Yhtälöllä $4x = e^x$ on kaksi ratkaisua. Etsi niille likiarvot Newtonin menetelmällä.

Vihje s. 158 →

Toinen vihje s. 173 →

Tehtävä 40. Tehtävässä 39 ratkaistiin yhtälö $4x = e^x$ kaksi ratkaisua Newtonin menetelmällä. Selvitä molempien ratkaisujen viisidesimaaliset likiarvot kiintopisteiteraation avulla.

Vihje s. 159 →

Toinen vihje s. 173 →

Tehtävä 41. Keplerin yhtälö

$$M = E - e \sin E,$$

on eräs perustavanlaatuinen yhtälö, joka liittyy taivaankappaleiden liikkeisiin. Kahden taivaankappaleen vuorovaikuttaessa liikkeiden laskemiseen liittyy suuret nimeltä **keskianomalia** M ja **eksentrinen anomalia** E . Lisäksi suureessa esiintyvä parametri e on niin sanottu radan **eksentrisyys** (ei siis Neperin luku tässä).

Selvitä Newtonin menetelmällä eksentrinen anomalia E , kun keskianomalia $M = 1,0$ ja eksentrisyys $e = 0,5$. Voiko ratkaisuja olla useampia?

Vihje s. 159 →

Toinen vihje s. 173 →

Tehtävä 42. Etsi Newtonin menetelmällä luvun 10 kuutiojuurelle likiarvo viiden desimaalin tarkkuudella. Käyttämäsi menetelmän ei tulisi käyttää muita laskutoimituksia kuin yhteen-, vähennys-, kerto- ja jakolaskua.

Vihje s. 159 →

Toinen vihje s. 173 →

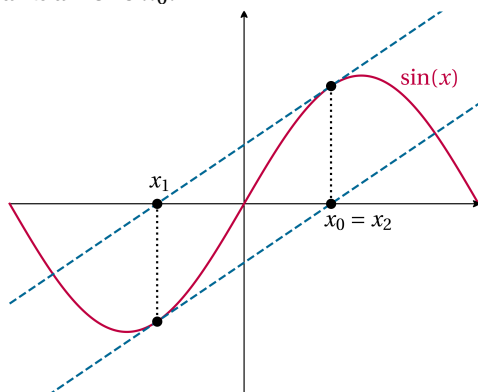
Tehtävä 43. Laadi Newtonin menetelmään perustuva algoritmi, joka laskee positiivisen reaaliluvun a käänteisluvulle $1/a$ likiarvon käyttäen vain yhteen-, vähennys- ja kertolaskua.

Laske menetelmällä luvun $1/\pi$ likiarvo. Kuinka monta kertolaskua tietokoneen täytyy laskea, jotta likiarvossa on 8 desimaalia oikein?

Vihje s. 159 →

Toinen vihje s. 173 →

Tehtävä 44. Tutkimme Newtonin menetelmää tapauksessa $f(x) = \sin(x)$. On olemassa alkuarvo $0 < x_0 < \frac{\pi}{2}$, jolle menetelmää iteroidessa käy niin, että $x_2 = x_0$. Selvitä likiarvo tälle alkuarvolle x_0 .



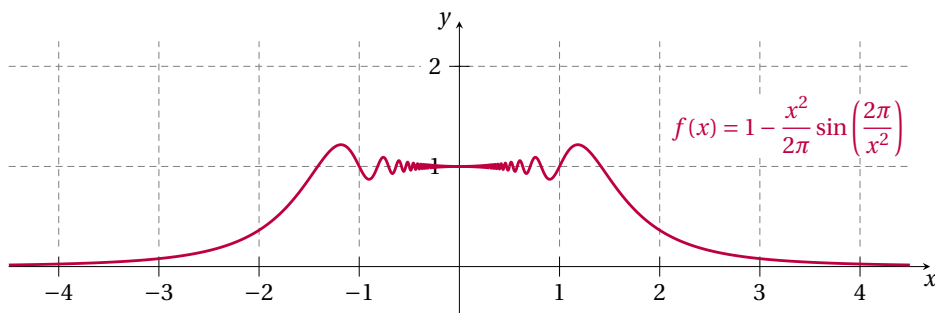
Vihje s. 159 →

Toinen vihje s. 174 →

Tehtävä 45. Newtonin iteraatio suppenee, mutta ei nollakohtaan. On mahdollista muodostaa derivoituva funktio, jolla Newtonin menetelmä suppenee lukuun, joka ei ole funktion nollakohta. Eräs tällainen esimerkki on

$$f(x) = \begin{cases} 1 - \frac{x^2}{2\pi} \sin\left(\frac{2\pi}{x^2}\right), & \text{kun } x \neq 0 \\ 1, & \text{kun } x = 0. \end{cases}$$

Osoita, että alkuarvolla $x_0 = 1$ Newtonin iteraatio suppenee lukuun, joka ei ole funktion f nollakohta. Oletetaan tässä tehtävässä, että laskut suoritetaan tarkoilla arvoilla.



Vihje s. 159 →

Toinen vihje s. 174 →

Tehtävä 46. Tutki kiintopisteiteraatiota tapauksessa $g(x) = x^2$. Selvitä ensin, miten alkupisteen x_0 valinta vaikuttaa menetelmän suppenemiseen. Tutki vielä derivaatan avulla, ovatko tämän funktion kiintopisteet vakaita tai epävakaita.

Vihje s. 160 →

Toinen vihje s. 175 →

Tehtävä 47. Edellisessä tehtävässä esiintynyt kiintopisteyhtälö $x^2 = x$ voidaan (ei-negatiivisilla x :n arvoilla) kirjoittaa myös muodossa $\sqrt{x} = x$. Selvitä myös funktion $g(x) = \sqrt{x}$ kiintopisteiden vakaus ja epävakaus.

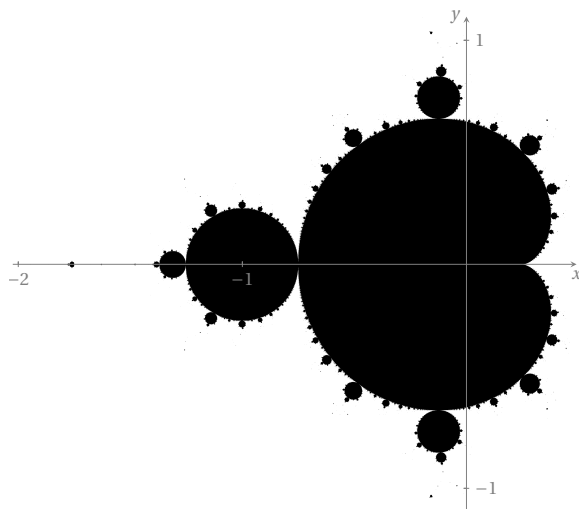
Vihje s. 160 →

Toinen vihje s. 175 →

Tehtävä 48. Mandelbrotin joukkoon kuuluvat sellaiset kompleksitason pisteet c , joilla lukujonon

$$\begin{cases} z_0 = c, \\ z_{n+1} = z_n^2 + c, \quad \text{kun } n = 0, 1, 2, \dots \end{cases}$$

jäsenet pysyvät jonkin ylärajan päässä origosta kaikilla $n \geq 0$. Joukkoon kuulumisen ehto voidaan muotoilla myös niin, että kun funktiota $g(z) = z^2 + c$ iteroidaan alkuarvolla $z_0 = c$, jonon jäsenten etäisyys origosta pysyy rajoitettuna.



Mandelbrotin joukko

Osalla Mandelbrotin joukkoon kuuluvista luvuista iteraatio päättyy sykliin, kuten luvulla -1 , jolla

$$\begin{aligned} z_0 &= c = -1 \\ z_1 &= z_0^2 + c = (-1)^2 - 1 = 0 \\ z_2 &= z_1^2 + c = 0^2 - 1 = -1 \\ z_3 &= 0, \end{aligned}$$

ja niin edelleen. Osalla taas iteraatio suppenee kiintopisteeseen, kuten luvulla $1/4$, jolla iteraatio lähestyy lukua $1/2$.

Tutki, millä luvun c reaalilla arvoilla funktiolla $g(x) = x^2 + c$ on vakaa kiintopiste.

Vihje s. 160 →

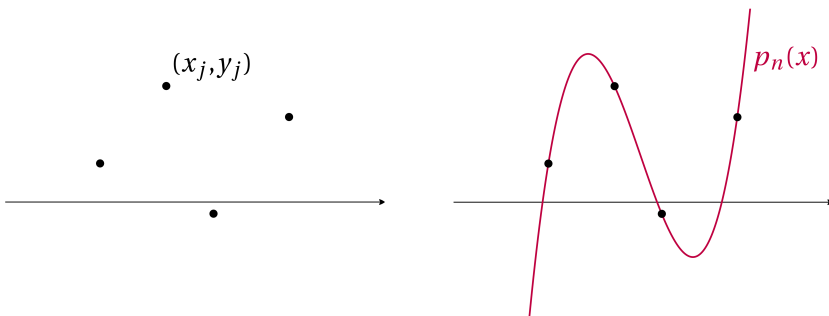
Toinen vihje s. 175 →

6 Lagrangen interpolaatio



Joseph-Louis Lagrange, 1736–1813

Tutustuimme aiemmin Taylorin polynomeihin, jotka tarjoavat luontevan tavan arvioida funktiota polynomifunktiolla minkä tahansa annetun pisteen läheisyydessä. Taylorin polynomit eivät suinkaan ole ainut tapa approksimoida funktioita polynomeilla, ja tutustumme tässä luvussa toiseen perusmenetelmään nimeltä **Lagrangen interpolaatio**. Tämän menetelmän tavoite on löytää polynomifunktio, jonka kuvaaja kulkee jonkin annetun pistejoukon kautta.



Kuva 22. Lagrangen interpolaatiossa sovitetaan polynomi annetun pistejoukon kautta.

Merkitään tällaista pistejoukkoa (x_j, y_j) , missä $j = 0, \dots, n$, eli pisteitä on $n + 1$ kappaletta. Haluaisimme löytää polynomin $p_n(x)$, jolle pätee

$$p_n(x_j) = y_j \quad \text{kaikilla } j = 0, \dots, n, \quad (18)$$

eli sovittaa polynomin p_n kuvaaja tason pisteiden (x_j, y_j) kautta. Tätä varten täytyy toki olettaa, että koordinaatit x_j ovat erisuuria, sillä funktiomme ei voi saada kahta eri arvoa samassa pisteessä. On luonnollista kysyä myös jatkokysymyksenä, mikä on yksinkertaisin mahdollinen polynomi $p_n(x)$ joka toteuttaa ehdot (18). Käytännössä tämä tulee tarkoittamaan sitä, että haluamme määrittää pienimmän mahdollisen asteen polynomille p_n . Aloitetaan tutkimaan ongelmaa yksinkertaisista tapauksista lähtien.

Mikäli pistejoukossamme on vain yksi piste (x_0, y_0) , yksinkertaisin ratkaisu löytyy valitsemalla polynomiksi vakiofunktio $p_0(x) = y_0$, jonka aste on nolla. Mikäli pisteitä on kaksi, eli (x_0, y_0) ja (x_1, y_1) , ei vakiofunktion sovittaminen välttämättä enää onnistu (jos $y_0 \neq y_1$). Sen sijaan ensimmäisen asteen polynomin sovittaminen aina onnistuu, sillä kahden pisteen kautta voi aina piirtää suoran. Tälle suoralle voi tarvittaessa kirjoittaa tarkan yhtälön

$$p_1(x) = (x - x_0) \frac{y_1 - y_0}{x_1 - x_0} + y_0,$$

joka on haluttu ensimmäisen asteen polynomi. Näistä kahdesta esimerkistä voi ehkä jo arvata, että tyypillisesti $n + 1$ pisteen kautta sovittamiseen tarvitaan ratkaisuksi polynomi astetta n . Näin todellakin käy, ja tämä on varsin luonnollinen lopputulos, sillä polynomi astetta n on muotoa

$$p_n(x) = a_n x^n + \dots + a_1 x + a_0,$$

eli sen määrittää yhteensä $n + 1$ säädettävää kerrointa $a_0, \dots, a_n$. Kertoimia on siis yhtä paljon kuin yhtälöitä kaavassa (18), mikä on luonnollinen tilanne yhtälöryhmän ratkaisemisen kannalta.

Osoittautuu myös, että mikäli yhtälöryhmällä (18) on enintään astetta n oleva polynomi $p_n(x)$, niin sen täytyy olla yksikäsitteinen. Jos nimittäin kaksi eri polynomia $p(x)$ ja $q(x)$ toteuttavat nämä samat yhtälöt, niin funktiolla $p(x) - q(x)$ olisi ainakin $n + 1$ nollakohtaa $x_0, \dots, x_n$. Polynomilla, jonka aste on enintään n , ei voi olla $n + 1$ nollakohtaa ellei polynomi ole vakiofunktio nolla. Siispä $p(x) - q(x) = 0$, eli $p = q$.

Tarvitsisi enää löytää yhtälöryhmälle (18) jokin ratkaisu, joka on enintään astetta n . Tämä ratkaisu voidaan aina löytää, ja sitä sanotaan pistejoukon (x_j, y_j) **Lagrangen interpolaatiopolynomiksi**. Tutkitaan nyt, miten tämä polynomi voidaan määrittää.

Hyvin konkreettinen menetelmä löytää Lagrangen interpolaatiopolynomi on käsitellä ensin helpompi tapaus, jossa arvoista y_j kaikki paitsi yksi on nolla. Sovitaan

esimerkiksi, että $y_j = 0$ aina, kuin $j \neq k$, missä k on jokin kiinnitetty indeksi joukosta $0, \dots, n$. Yksinkertaistetaan vielä sopimalla, että $y_k = 1$. Haluaisimme siis määrittää polynomin L_k , jolla on nollakohta pisteissä $x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n$, ja joka saa pisteessä x_k arvon 1. Koska L_k on enintään astetta n , olemme löytäneet kaikki sen mahdolliset n juurta. Siis voimme kirjoittaa polynomin L_k muodossa

$$L_k(x) = C_k(x - x_0) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n) = C_k \prod_{j \neq k} (x - x_j),$$

missä $\prod_{j \neq k} (x - x_j)$ on lyhennysmerkintä yllä olevalle tulolle ja C_k on vakiokerroin, jonka voimme vielä määrittää. Tämä vakio selviää ehdosta $L_k(x_k) = 1$, josta saadaan yhtälö

$$1 = L_k(x_k) = C_k \prod_{j \neq k} (x_k - x_j), \quad \text{josta ratkaistuna} \quad C_k = \frac{1}{\prod_{j \neq k} (x_k - x_j)}.$$

Siis polynomille L_k saadaan lopullinen kaava

$$L_k(x) = \frac{\prod_{j \neq k} (x - x_j)}{\prod_{j \neq k} (x_k - x_j)}.$$

Palataan nyt yleiseen tapaukseen, jossa arvot y_j ovat mielivaltaisia. Nähdään kuitenkin, että tämän ongelman ratkaisu saadaan kaavasta

$$p_n(x) = y_0 L_0(x) + y_1 L_1(x) + \cdots + y_n L_n(x),$$

sillä sijoittamalla saadaan $L(x_k) = 0 + 0 + \cdots + y_k L_k(x_k) + 0 + \cdots + 0 = y_k$. Siis yleisen Lagrangen interpolaatiopolynomin eli yhtälöryhmän (18) yksikäsitteisen ratkaisun astetta $\leq n$ voi kirjoittaa muodossa

$$p_n(x) = \sum_{k=0}^n y_k L_k(x) = \sum_{k=0}^n y_k \frac{\prod_{j \neq k} (x - x_j)}{\prod_{j \neq k} (x_k - x_j)}. \quad (19)$$

Esimerkki. Paraabeli kolmen pisteen kautta

Sovelletaan kaavaa (19) tilanteeseen, jossa halutaan yhdistää pisteet (a, A) , (b, B) ja (c, C) toisen asteen polynomin kuvaajalla. Lagrangen polynomiksi p_2 saadaan

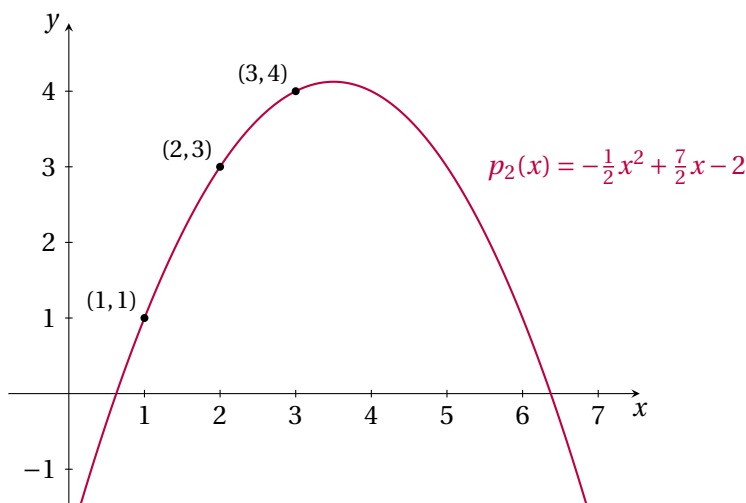
$$p_2(x) = A \cdot \frac{(x-b)(x-c)}{(a-b)(a-c)} + B \cdot \frac{(x-a)(x-c)}{(b-a)(b-c)} + C \cdot \frac{(x-a)(x-b)}{(c-a)(c-b)}.$$

Kaavan toimintaa voi havainnollistaa sijoittamalla $x = a$, jolloin saadaan

$$p_2(a) = A \cdot \frac{(a-b)(a-c)}{(a-b)(a-c)} + B \cdot 0 + C \cdot 0 = A \cdot 1 = A.$$

Valitaan pisteiksi esimerkiksi $(1, 1)$, $(2, 3)$ ja $(3, 4)$, jolloin polynomiksi saadaan hieman sieventämällä

$$\begin{aligned} p_2(x) &= 1 \cdot \frac{(x-2)(x-3)}{(1-2)(1-3)} + 3 \cdot \frac{(x-1)(x-3)}{(2-1)(2-3)} + 4 \cdot \frac{(x-1)(x-2)}{(3-1)(3-2)} \\ &= 1 \cdot \frac{x^2 - 5x + 6}{2} + 3 \cdot \frac{x^2 - 4x + 3}{-1} + 4 \cdot \frac{x^2 - 3x + 2}{2} \\ &= -\frac{1}{2}x^2 + \frac{7}{2}x - 2. \end{aligned}$$



6.1 Lagrangen polynomin virhe

Pohditaan nyt, miten hyvin Lagrangen interpolaatiota voidaan soveltaa jonkin annetun funktion approksimoimiseen. Olkoon funktio $f(x)$ määritelty välillä $[a, b]$, ja valitaan väliltä $[a, b]$ yhteensä $n + 1$ kappaletta pisteitä $x_0, \dots, x_n$. Valitsemalla Lagrangen interpolaatiossa $y_j = f(x_j)$ voidaan nyt funktion kuvaajan pisteiden $(x_j, f(x_j))$ kautta sovittaa polynomi $p(x)$ astetta n kuten edellä, ja tämä polynomi saa erityisesti samat arvot kuin funktio f valitsemissamme pisteissä x_j .

Kysymys kuuluu, kuinka hyvin polynomi p approksimoi funktion f arvoja muissa pisteissä kuin kiinnittämissämme pisteissä x_j . Yleisesti ei ole mitään taetta, että funktion f kuvaaja ei mutkittelisi kaukana polynomista pisteiden x_j ulkopuolella, mutta olettamalla hieman lisää funktion f derivoituvuudesta voidaan todistaa seuraava tulos.

Lagrangen interpolaation virhearvio

Mikäli funktio f on $n + 1$ kertaa jatkuvasti derivoituva välillä $[a, b]$, on jokaisella pisteellä x olemassa piste $\xi = \xi(x)$ välillä $]a, b[$ siten, että

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \pi_{n+1}(x), \quad (20)$$

missä $f^{(n+1)}$ tarkoittaa funktion asteen $n + 1$ derivaattaa, ja

$$\pi_{n+1}(x) = (x - x_0) \cdots (x - x_n).$$

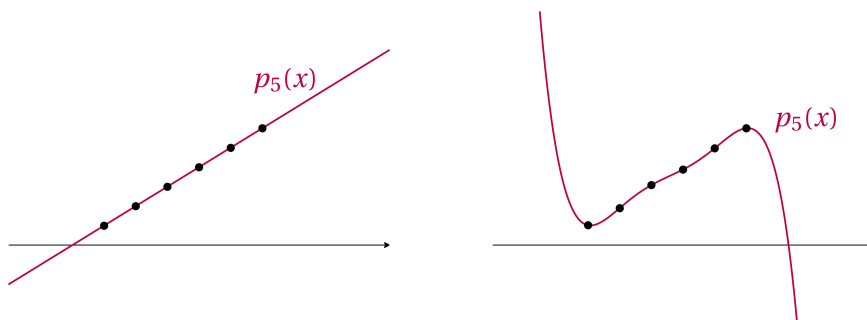
Todistamme tämän lauseen hetken päästä, mutta haluamme ensin kiinnittää huomiota siihen, että yhtälö (20) antaa meille arvion funktion f ja polynomin p_n arvojen etäisyydestä. Nimittäin jos $M_{n+1} = \max_{\xi \in [a, b]} |f^{(n+1)}(\xi)|$ merkitsee kertaluvun $n + 1$ derivaatan $f^{(n+1)}$ maksimiarvoa välillä $[a, b]$, niin tällöin saadaan epäyhtälö

$$|f(x) - p_n(x)| \leq \frac{|f^{(n+1)}(\xi)|}{(n+1)!} |\pi_{n+1}(x)| \leq \frac{M_{n+1}(b-a)^{n+1}}{(n+1)!}. \quad (21)$$

Tässä viimeinen epäyhtälö perustuu myös arvioon $|x - x_j| \leq b - a$, eli siihen, että kahden välin $[a, b]$ pisteen etäisyys on enintään välin pituus, joka voidaan tehdä jokaiselle funktiossa π_{n+1} esiintyvälle tekijälle.

Vaikka välin $[a, b]$ pituus $b - a$ olisi iso, huomataan, että arviossa (21) esiintyvä eksponentiaalinen tekijä $(b - a)^{n+1}$ kasvaa silti huomattavasti hitaammin kuin kertomafunktio $(n + 1)!$ nimittäjässä. Arvio (21) kertoo siis erityisesti sen, että mikäli funktion f korkeamman asteen derivaatat eivät paisu erityisen suuriksi, niin Lagrangen interpolaatiopolynomi antaa hyvin tarkkoja arvioita, kun aste n on valittu riittävän suureksi. Näin tapahtuu esimerkiksi trigonometrisilla funktioilla $\sin(x)$ ja $\cos(x)$, joiden kaikkien asteiden derivaatat pysyvät samoissa rajoissa kuin funktiot itse.

Huom! Yllä esitetty virhekaava antaa hyviä arvioita yksinomaan silloin, kun annetun pistejoukon kautta kulkee funktio f , jonka asteen $n + 1$ derivaatta $f^{(n+1)}$ ei saa liian suuria arvoja. Tämä on tilanne, joka ei käytännön sovelluksissa suinkaan aina toteudu. Nimittäin pienikin muutos interpoloitavassa datassa voi johtaa suuriin muutoksiin asteen $n + 1$ derivaatan arvioissa, ja aiheuttaa viljejä heilahteluja Lagrangen interpolaatiopolynomissa varsinkin, kun ekstrapoloidaan alkuperäisen datan ulkopuolelle.



Kuva 23. Pienet muutokset lähtöarvoissa voivat johtaa suuriin muutoksiin Lagrangen interpolaatiopolynomin arvoissa.

Esimerkiksi reaalimaailman sovelluksissa mitattavaan dataan liittyy aina tiettyjä mittausrvirheitä, jonka vuoksi Lagrangen interpolaatio ei ole näissä tilanteissa suositeltava sovitustmenetelmä. Siksi käytännön sovelluksissa turvaudutaan usein muihin approksimointimenetelmiin, kuten ns. **pienimmän neliösumman menetelmään**.

Virhekaavan todistus. Todistetaan nyt kaava (20). Todistus ei ole kovin yksinkertainen, sillä se perustuu Luvussa 3 esittämäämme Rollen lauseeseen. Rollen lause sanoo, että jos jollakin derivoituvalla funktiolla F on kaksi nollakohtaa, niin sen derivaatalla on nollakohta näiden välissä.

Samalla idealla voidaan saada myös tietoa korkeamman asteen derivaatoista mikäli nollakohtia on enemmän. Jos funktiolla F on esimerkiksi kolme nollakohtaa a , b ja c , niin sen derivaatalla F' on ainakin kaksi nollakohtaa, yksi molemmilla väleillä $]a, b[$ ja $]b, c[$. Soveltamalla Rollen lausetta uudestaan funktioon F' näissä nollakohdissa huomataan, että myös toisella derivaatalla F'' täytyy olla nollakohta välillä $]a, b[$.

Induktiolla voidaan tästä todistaa yleinen **Rollen lause korkeamman asteen derivaatoille**, joka sanoo, että jos $n + 1$ kertaa derivoituvalla funktiolla F on $n + 2$ nollakohtaa, niin sen $n + 1$ -kertaisella derivaatalla $F^{(n+1)}$ on ainakin yksi nollakohta näiden välissä.

Palataan nyt kaavan (20) todistamiseen. Loppuosa todistuksesta tulee vaatimaan vielä yhden tempun. Haluamme todistaa yhtälön (20) jokaiselle x , joten valitaan jokin x väliltä $[a, b]$, ja ajatellaan sitä kiinnitettynä arvona. Jos $x = x_j$ jollakin j , niin molemmat puolet yhtälöstä (20) ovat selvästi nollija, eli voimme olettaa $x \neq x_j$ kaikilla j . Nyt niksi on tarkastella sopivaa lauseketta, joka on funktio $F(t)$ muotoa

$$F(t) = f(t) - p_n(t) - \frac{f(x) - p_n(x)}{\pi_{n+1}(x)} \pi_{n+1}(t).$$

Otimme tässä käyttöön uuden muuttujan t , koska x on kiinnitetty luku.

Nyt huomataan, että funktiolla $F(t)$ todellakin on $n + 2$ nollakohtaa välillä $[a, b]$, sillä se on nolla $n + 1$ pisteessä $x_0, \dots, x_n$, ja nähdään myös helposti laskemalla, että $F(x) = 0$. Erityisesti Rollen lause sanoo, että $n + 1$ -kertaisella derivaatalla $F^{(n+1)}(t)$ on nollakohta välillä $[a, b]$.

Toisaalta funktiota F ei ole vaikea derivoida: Polynomi $p_n(x)$ on astetta n , joten sen $n + 1$ -kertainen derivaatta on nolla. Polynomi π_{n+1} taas on astetta $n + 1$, joten sen $n + 1$ -kertainen derivaatta on vakio, joka määräytyy suurimman asteen termin kertoimen perusteella. Kertomalla sulut auki huomataan, että

$$\pi_{n+1}(t) = (t - x_0) \cdots (t - x_n) = t^{n+1} + [\text{matalamman asteen termit}],$$

joten potenssifunktiota t^{n+1} toistuvasti derivoimalla saadaan

$$\pi_{n+1}^{(n+1)}(t) = (n + 1) \cdot n \cdot \cdots \cdot 2 \cdot 1 = (n + 1)!.$$

Siispä voidaan lopulta laskea, että

$$F^{(n+1)}(t) = f^{(n+1)}(t) - \frac{f(x) - p_n(x)}{\pi_{n+1}(x)}(n + 1)!.$$

Rollen lauseen perusteella tällä funktiolla oli nollakohta jossain pisteessä $t = \xi$, joten tässä pisteessä pätee, että

$$0 = f^{(n+1)}(\xi) - \frac{f(x) - p_n(x)}{\pi_{n+1}(x)}(n + 1)!,$$

mistä seuraa väittämämme yhtälö (20), kun $x \neq x_j$. Väite on todistettu. $\square$

Kaava (20) ja siitä johdettu arvio (21) saattavat vaikuttaa hieman mutkikkailta, mutta niillä on mielenkiintoisia seurauksia.

Palautetaan mieleen, että Lagrangen interpolaatiopolynomi p_n on yksikäsitteinen polynomi astetta $\leq n$, joka leikkaa funktion kuvaajaa $n + 1$ pisteessä. Siispä arvio (21) pätee itse asiassa **mille tahansa** polynomille astetta $\leq n$, jonka tiedetään leikkaavan funktion f kuvaajaa jossain $n + 1$ pisteessä (koska tämän polynomin täytyy olla interpolaation antama polynomi p_n).

Tämä huomio on toisinaan hyödyllinen arvioimaan polynomin erotusta johonkin tunnettuun funktioon, ja tarvitsemme yllä esiteltyjä virhearvioita etenkin seuraavassa luvussa, jossa perehdymme yhteen polynomiapproksimaation keskeisimmistä sovelluksista, numeeriseen integrointiin.

Harjoitustehtäviä

Tehtävä 49. Etsi Lagrangen interpolaatiopolynomi pisteille $(-1,1)$, $(1,1)$, ja $(2,4)$ kaavasta (19) ja sievennä se.

Vihje s. 160 →

Toinen vihje s. 175 →

Tehtävä 50. Sekä Lagrangen interpolaatio että Taylorin polynomi ovat tapoja approksimoida tunnettua funktiota polynomifunktioilla. Näiden lisäksi on muitakin tapoja, kuten Hermiten ja Chebyshevin polynomit.

Verrataan nyt Lagrangen ja Taylorin polynomien muodostamaa approksimaatiota funktiolle $f(x) = \cos x$. Muodosta ensin toisen asteen Lagrangen polynomi, joka saa samat arvot kuin funktio f pisteissä $-\pi/2$, 0 , ja $\pi/2$. Vertaa tulosta tehtävässä 12 laskettuun toisen asteen Taylorin polynomiin, joka on kehitetty pisteessä $x = 0$. Millaiseen approksimointiin eri polynomit sopivat?

Vihje s. 160 →

Toinen vihje s. 175 →

Tehtävä 51. Oletetaan, että tunnemme pisteiden (x_j, y_j) , $j = 1, \dots, n+1$ kautta kulkevan interpolaatiopolynomin p_n . Haluaisimme lisätä joukkoon vielä uuden pisteen (x_0, y_0) ja muodostaa uuden polynomin p_{n+1} tämän yhtä suuremman pistejoukon kautta. Suoraan kaavaan (19) sijoittaminen tuottaa kuitenkin vaivalloisen sieventämisen, joten keksi oikotie, joka hyödyntää polynomia p_n .

Vihje s. 160 →

Toinen vihje s. 175 →

Tehtävä 52. Tutkitaan funktiota $f(x) = \sin(x)$ välillä $[-1, 1]$. Etsi jokin riittävä pisteiden lukumäärä n niin, että Lagrangen interpolaatiopolynomin $p_n(x)$ absoluuttinen virhe $|f(x) - p_n(x)|$, on suuruudeltaan enintään 10^{-5} koko välillä.

Vihje s. 160 →

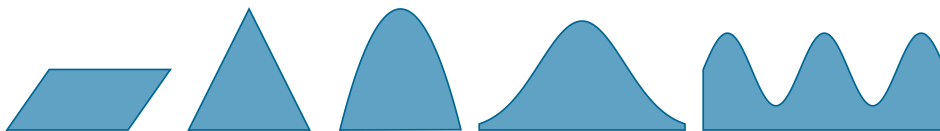
Toinen vihje s. 175 →

Tehtävä 53. Oletetaan, että funktiolla f on välillä $[a, b]$ seuraava ominaisuus: On olemassa luku N siten, että valittiinpa väliltä $[a, b]$ kuinka monta interpolaatiopistettä x_j hyvänsä, niin Lagrangen interpolaatiopolynomin aste on aina enintään N . Osoita, että funktio f on koko välillä $[a, b]$ yhtä suuri kuin jokin polynomi, joka on enintään astetta N .

Vihje s. 160 →

Toinen vihje s. 176 →

7 Numeerinen integrointi



Pinta-alojen laskeminen on suhteellisen yksinkertaista niin kauan kuin käsitellään monikulmioita, mutta kaarevien muotojen alat lasketaan yleensä määrättyjen integraalien $\int_a^b f(x) dx$ avulla.

Tässä luvussa pyrimme ymmärtämään, miten määrättyjen integraalien arvoja voidaan laskea numeerisesti. Tyypillisesti opimme ensin laskemaan integraaleja määrittämällä integroitavalle funktiolle integraalifunktio, jonka avulla määrätylle integraalille voidaan laskea tarkka arvo käyttäen **analyysin peruslauseen** muotoa

$$\int_a^b f(x) dx = F(b) - F(a), \quad \text{missä } F'(x) = f(x).$$

Toisaalta monilla funktioilla ei ole ollenkaan yksinkertaista integraalifunktiota, tai sellaisen laskeminen ei ole käytännöllistä. Funktio e^{x^2} on perusesimerkki funktiosta, jonka integraalifunktiota ei voida esittää tavallisten funktioiden avulla. Tällaisissa tapauksissa määrättylläkään integraalilla ei yleensä ole suljettua muotoa.

Eräs tällainen integraali on

$$\int_0^1 e^{x^2} dx \approx 1,46265,$$

jonka kanssa joudumme tyytymään likiarvoon. Monessa sovelluksessa emme muuta tarvitsekaan. Määrätyn integraalin likiarvon laskemista kutsutaan **numeeriseksi integroinniksi**. Hyvän numeerisen integrointimenetelmän kriteereitä ovat laskennan tarkkuus ja keveys: haluamme riittävän tarkan likiarvon lyhyessä laskenta-ajassa.

Yleinen kysymyksemme on siis laskea likiarvo annetun funktion f määrätylle integraalille

$$\int_a^b f(x) dx, \tag{21}$$

jonkin annetun reaaliakselin välin $[a, b]$ yli. Integroimista varten oletamme alkuun ainakin, että f on jatkuva.

Tyypillisin perusidea numeerisessa integroinnissa on korvata annettu funktio f jollain yksinkertaisemmalla funktiolla, joka pyritään valitsemaan niin, että se approksimoi funktiota f mahdollisimman hyvin. Eräs luonnollinen valinta tähän tehtävään ovat polynomifunktiot, sillä polynomien integraaleja on helppo laskea. Kuten olemme aiemmissa luvuissa havainneet, on olemassa luontevia tapoja approksimoida funktioita polynomeilla, kuten Taylorin ja Lagrangen polynomit.

Polynomeilla approksimoitaessa olemme myös huomanneet, että polynomin asteen kasvattaminen yleensä parantaa arvioita. Näin käy myös numeerisessa integroinnissa, mutta aloitetaan tarkastelu kuitenkin yksinkertaisimmasta tapauksesta.

7.1 Suorakaidesääntö

Yksinkertaisin valinta polynomiapproksimaatioon on tietysti vakiopolynomi, eli astetta 0 oleva vakiofunktio $p_0(x) = c$. Mikäli annettua funktiota f halutaan approksimoida vakiofunktioilla välillä $[a, b]$, on luontevaa asettaa täksi vakioksi jokin funktion arvoista, esimerkiksi funktion f arvo välin keskipisteessä $\frac{a+b}{2}$. Tästä saadaan seuraavanlainen arvio

$$\int_a^b f(x) dx \approx \int_a^b f\left(\frac{a+b}{2}\right) dx = (b-a)f\left(\frac{a+b}{2}\right). \quad (22)$$

Tämä arvio ei välttämättä ole hirveän hyvä, mutta se on ainakin helppo laskea. Samalla idealla voidaan kuitenkin johtaa myös tarkempia arvioita, jakamalla yksinkertaisesti väli $[a, b]$ ensin pienempiin osaväleihin.

Aloitetaan valitsemalla osavälien lukumäärä n , ja jakamalla väli $[a, b]$ tasaisesti samansuuruisiin osaväleihin. Jokaisen osavälin pituudeksi tulee siis

$$\Delta x = \frac{b-a}{n}.$$

Jos osavälejä merkitään $I_1, \dots, I_n$, niin niiden tarkat lausekkeet saadaan kaavoista

$$\begin{aligned} I_1 &= [a, a + \Delta x] \\ I_2 &= [a + \Delta x, a + 2\Delta x] \\ &\vdots \\ I_n &= [a + (n-1)\Delta x, b]. \end{aligned}$$

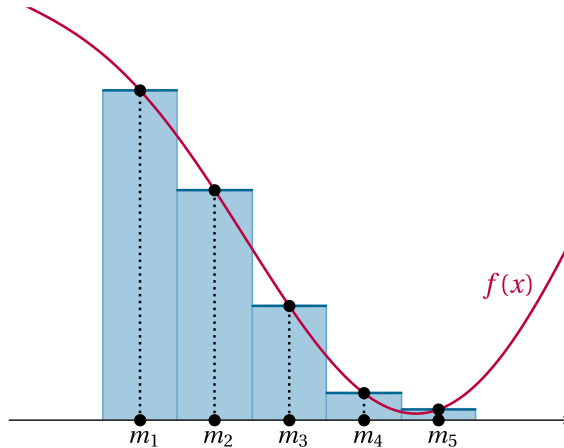
Tarkkaillaan myös jokaisen osavälin I_j keskipistettä m_j , jolle saadaan kaavaksi $m_j = a + \left(j - \frac{1}{2}\right)\Delta x$.

Mitä suurempi luku n valitaan, sitä pienempiä osavälit I_j ovat. Kun funktio f on oletettu jatkuvaksi ja välit valitaan riittävän pieniksi, niin funktion f arvot jokaisen yksittäisen pienen välin I_j sisällä eivät poikkea kovin paljon toisistaan. Siispä mitä pienempi osaväli I_j , sitä pienempi virhe vakioapproksimaatiossa $f(x) \approx f(m_j)$ tapahtuu, kun x on valittu väliltä I_j .

Pilkotaan nyt funktion f määrätty integraali osaväleihin, ja sovelletaan jokaisella osavälillä arviota (22). Tämä antaa tulokseksi arvion

$$\int_a^b f(x) dx \approx \Delta x \sum_{j=1}^n f(m_j).$$

Tätä arviota sanotaan **suorakaidesäännöksi**, ja sen antama arvio suppenee kohti määrätyn integraalin todellista arvoa, kun $n \rightarrow \infty$. Suorakaidesääntö saa nimensä sen geometrisesta tulkinnasta, sillä jos funktion määrätyn integraalin tulkitsee sen kuvaajan alle jäävänä pinta-alana, niin suorakaidesääntö perustuu pinta-alan arvioimiseen kapeilla suorakaiteilla. Suorakaidesäännön antaman arvion suppe-

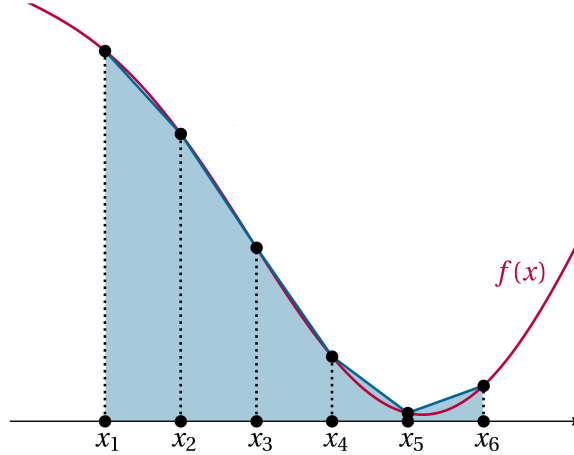


Kuva 24. Suorakaidesääntö viidellä osavälillä.

nemisnopeus riippuu funktion f käyttäytymisestä, sillä funktion heilahtelevuuden pienemmällä skaaloilla voi pahentaa arviointivirhettä. Suorakaidesääntö on melko helppo toteuttaa, sillä se vaatii ainoastaan funktion f arvojen laskemista osavälien keskipisteissä. Se osoittautuu kuitenkin hieman epäkäytännölliseksi menetelmäksi, sillä korkeamman asteen menetelmät tuottavat huomattavasti parempia arvioita. Palaamme tarkkaan virhearviointiin myöhemmin, mutta käsitellään ensin lisää menetelmiä.

7.2 Puolisuunnikassääntö

Yksinkertainen tapa pyrkiä parantamaan suorakaidesäännön arviota on kasvattaa approksimoivan funktion astetta yhdellä, eli korvata vakiofunktio lineaarisella funktiolla. Lineaarisen funktion kuvaaja välillä $[a, b]$ on jana, joten geometrisesti tämä menettely vastaa suorakaidesäännön suorakulmioiden korvaamista puolisuunnikkailla, mistä menetelmä saa nimensä.



Kuva 25. Puolisuunnikassääntö viidellä osavälillä.

Haluaisimme seuraavaksi etsiä lineaarisen funktion, joka approksimoi funktiota f välillä $[a, b]$. Eräs luonnollinen tapa tehdä tämä on laskea ensin funktion arvo välin päätepisteissä a ja b . Päätepisteissä kuvaajan pisteiden $(a, f(a))$ ja $(b, f(b))$ kautta kulkee jana, jonka lausekkeeksi voidaan laskea $p_1(x) = \frac{f(b)-f(a)}{b-a}(x-a) + f(a)$. Tästä saadaan pienellä laskulla arvioksi

$$\int_a^b f(x) dx \approx \int_a^b p_1(x) dx = \frac{b-a}{2} (f(a) + f(b)). \quad (23)$$

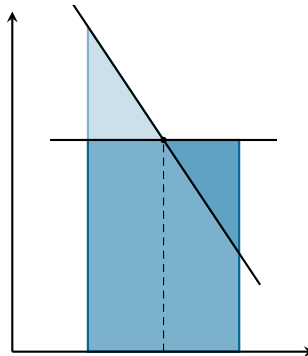
Kuten suorakaidesäännössä, tätä arviota voidaan parantaa toteuttamalla se välin $[a, b]$ pienissä osaväleissa koko välin sijaan. Merkitään näitä osaväleja taas I_j , ja olkoon $x_j = a + (j-1)\Delta x$ jokaisen osavälin I_j vasen päätepiste, jolloin I_j :n oikea päätepiste on x_{j+1} . Erityisesti $x_1 = a$ ja $x_{n+1} = b$. Nyt soveltamalla arviota (23) jokaisen osavälin sisällä saadaan

$$\begin{aligned} \int_a^b f(x) dx &\approx \Delta x \left(\frac{f(x_1) + f(x_2)}{2} + \frac{f(x_2) + f(x_3)}{2} + \dots + \frac{f(x_n) + f(x_{n+1})}{2} \right) \\ &= \Delta x \left(\frac{1}{2} f(x_1) + f(x_2) + f(x_3) + \dots + f(x_n) + \frac{1}{2} f(x_{n+1}) \right) \end{aligned}$$

$$= \Delta x \left(\frac{f(a)}{2} + \sum_{j=2}^n f(x_j) + \frac{f(b)}{2} \right).$$

Verrattuna suorakaidesääntöön, tulee puolisuunnikassäännössä laskea funktion arvo ainoastaan yhdessä ylimääräisessä pisteessä, sillä osavälien päätepisteitä on yhteensä $n + 1$ kappaletta kun taas keskipisteitä on n kpl. Vaikka puolisuunnikassääntö on yhtä polynomin astetta korkeampi menetelmä kuin suorakaidesääntö, ei se kuitenkaan tarjoa huomattavasti parempaa arviota integraalille. Tämä perustuu seuraavaan ideaan.

Palataan vielä suorakaidesäännön vakiofunktioon $p_0(x) = f\left(\frac{a+b}{2}\right)$. Huomataan, että tällä vakiofunktioilla on sama integraali kuin millä tahansa lineaarisella funktiolla, joka saa saman arvon kuin p_0 välin keskipisteessä $m = \frac{a+b}{2}$, eli millä tahansa funktiolla muotoa $q(x) = C(x - m) + f(m)$, missä C on vapaavalintainen kulma-kerroin. Tälle seikalle on luonteva geometrinen selitys, joka näkyy kuvassa 26 tapahtuvana vakiofunktion yli ja ali menevien alueiden kumoutumisena. Tämä kumoutuminen perustuu erityisesti siihen, että ensimmäisen asteen polynomitermi x on pariton funktio.



Kuva 26. Vakiofunktioilla ja lineaarisella funktiolla on sama integraali symmetrisellä välillä funktioiden leikkauspisteen ympärillä.

Jossain mielessä suorakaidesääntö siis vastaa myös ensimmäisen asteen polynomiapproksimaatiota, sillä se tarjoaa saman arvion kuin jos vakiofunktion sijaan approksimaatioon käyttäisi lineaarista funktiota, joka saa saman arvon kuin f välin keskipisteessä ja missä tahansa toisessa vapaavalintaisessa pisteessä. Tämä antaa toki hitusen eri lopputuloksen kuin jos kyseisen lineaarisen funktion sovitaisi välin päätepisteiden suhteen, mutta virhearvion kannalta menetelmissä ei tule olemaan huomattavaa eroa.

Vaikka asteen kasvattaminen ei tällä kertaa auttanut, tapahtuu huomattava parannus toisen asteen approksimaatioon siirryttäessä.

7.3 Simpsonin sääntö

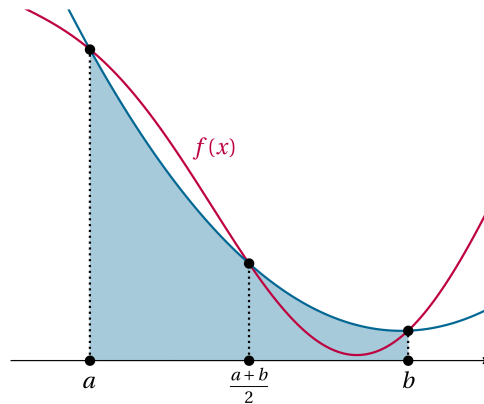
Luonnollisesti seuraava idea on approksimoida funktiota f toisen asteen polynomilla, eli arvioida sen kuvaajaa paraabelilla. Tämä lähestymistapa tunnetaan **Simpsonin sääntönä**, ja se on saanut nimensä Thomas Simpsonilta (1710–1761). Itse asiassa Simpson oppi menetelmän Newtonilta, ja saman menetelmän julkaisi jo Johannes Kepler vuonna 1615. Matemaattisten tulosten nimeämisen hienoisesta epäjohdonmukaisuudesta kertoo myös se, että Newtonin menetelmän nykyään käytössä olevan muodon kehitti puolestaan Simpson.



Thomas Simpson (1710–1761)

Paraabelin sovittamiseen täytyy valita kuvaajalta kolme pistettä, ja hyvin luonnollinen valinta tähän ovat välin päätepisteet ja keskipiste. Yksittäisellä välillä $[a, b]$ voidaan sovittaa Lagrangen interpolaatiolla toisen asteen polynomi $p_2(x)$, jolle pätee $p_2(a) = f(a)$, $p_2(b) = f(b)$, ja $p_2\left(\frac{a+b}{2}\right) = f\left(\frac{a+b}{2}\right)$. Tälle polynomille voidaan tarvittaessa laskea kaava lausekkeesta (19). Säästämme lukijan mekaaniselta laskulta, ja paljastamme, että tämä tuottaa arvioksi

$$\int_a^b f(x) dx \approx \int_a^b p_2(x) dx = \frac{b-a}{6} \left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right). \quad (24)$$



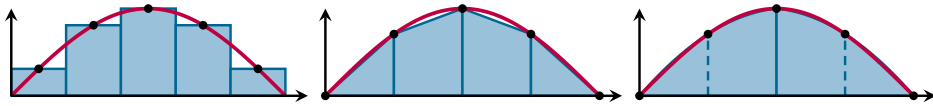
Kuva 27. Simpsonin säännössä funktion integraalia verrataan sopivasti sovitetun paraabelin integraaliin.

Tätä voidaan halutessamme soveltaa edelleen pienemmillä osaväleillä $I_j = [x_j, x_{j+1}]$, joiden keskipisteitä merkitsemme m_j , ja saadaan arvio

$$\begin{aligned} \int_a^b f(x) dx &\approx \frac{\Delta x}{6} [f(x_1) + 4f(m_1) + 2f(x_2) + 4f(m_2) + 2f(x_3) + 4f(m_3) + \\ &\quad + 2f(x_4) + \cdots + 4f(m_n) + f(x_{n+1})] \\ &= \frac{\Delta x}{6} \left[f(x_1) + 4 \sum_{j=1}^n f(m_j) + 2 \sum_{j=2}^n f(x_j) + f(x_{n+1}) \right]. \end{aligned}$$

Osoittautuu, että Simpsonin säännön tuottama arvio on huomattavasti parempi kuin suorakaidesäännössä ja puolisuunnikassäännössä. Tarkkuutta arvioitaessa tasapuolista on vertailla menetelmien versioita, jotka laskevat funktion arvon yhtä monessa pisteessä, sillä juuri funktion arvon laskeminen vie tyypillisesti eniten laskenta-aikaa.

Esimerkki. Testataan suorakaidesäännön, puolisuunnikassäännön ja Simpsonin säännön tarkkuutta laskemalla määrätty integraali $I = \int_0^\pi \sin x dx$ viiden funktion arvon avulla.



Kuva 28. Funktion $f(x) = \sin x$ välin $[0, \pi]$ määrätyn integraalin approksimointi suorakaidesäännöllä, puolisuunnikassäännöllä ja Simpsonin säännöllä.

Integraalin tarkka arvo on $I = 2$, joten menetelmien tarkkuutta on helppo vertailla. Menetelmät antavat seuraavat arvot:

Suorakaidesääntö:

$$\begin{aligned} I &\approx \Delta x \cdot (f(m_1) + f(m_2) + f(m_3) + f(m_4) + f(m_5)) \\ &= \frac{\pi}{5} \cdot \left(\sin \frac{\pi}{10} + \sin \frac{3\pi}{10} + \sin \frac{5\pi}{10} + \sin \frac{7\pi}{10} + \sin \frac{9\pi}{10} \right) \approx 2,03328. \end{aligned}$$

Puolisuunnikassääntö:

$$\begin{aligned} I &\approx \Delta x \cdot \left(\frac{1}{2} f(x_1) + f(x_2) + f(x_3) + f(x_4) + \frac{1}{2} f(x_5) \right) \\ &= \frac{\pi}{4} \cdot \left(\frac{1}{2} \sin(0) + \sin \frac{\pi}{4} + \sin \frac{\pi}{2} + \sin \frac{3\pi}{4} + \sin \pi \right) \approx 1,89612. \end{aligned}$$

Simpsonin sääntö:

$$\begin{aligned} I &\approx \frac{\Delta x}{6} \cdot (f(x_1) + 4f(m_1) + 2f(x_2) + 4f(m_2) + f(x_3)) \\ &= \frac{\pi/2}{6} \cdot \left(\sin(0) + 4 \sin \frac{\pi}{4} + 2 \sin \frac{\pi}{2} + 4 \sin \frac{3\pi}{4} + \sin \pi \right) \approx 2,00456. \end{aligned}$$

Tälle integraalille suorakaidesääntö antaa jopa paremman arvion kuin puolisuunnikassääntö, mutta Simpsonin sääntö on selvästi tarkin.

Tyypillisesti numeerista derivaattaa lasketaan tilanteessa, jossa tarkka arvo ei ole tiedossa. Tällöin on tärkeää pystyä arvioimaan menetelmän tarkkuutta teoreettisesti, jotta osataan valita menetelmä, joka antaa vastauksen tarvittavalla tarkkuudella.

7.4 Numeerisen integroinnin virhearviot

Haluaisimme esittää edellä esitettyjen numeeristen integrointimenetelmien virheen jollain tavalla, joka riippuisi funktion f käyttäytymisestä. Esittämämme virhearviot tulevat riippumaan funktiomme korkeamman asteen derivaatoista, joten oletamme tästä lähtien, että funktio f on derivoituva riittävän monta kertaa.

Tässä kohdassa Lagrangen interpolaatiopolynomeille todistamamme arvio (21) on hyödyksi, sillä edellä mainittujen numeeristen integroimismenetelmien perusidea oli juurikin sovittaa funktion kuvaajan pisteiden kautta sopivan asteinen polynomi.

Esimerkiksi puolisuunnikassäännön tapauksessa yksittäisellä välillä $[a, b]$ voidaan virheen itseisarvolle laskea arvio

$$\begin{aligned} \left| \int_a^b f(x) dx - \frac{(b-a)}{2} (f(a) + f(b)) \right| &= \left| \int_a^b f(x) dx - \int_a^b p_1(x) dx \right| \\ &= \left| \int_a^b (f(x) - p_1(x)) dx \right| \\ &\leq \int_a^b |f(x) - p_1(x)| dx \\ &\leq \int_a^b \frac{M_2(b-a)^2}{3!} dx \\ &= \frac{M_2}{6} (b-a)^3, \end{aligned}$$

missä sovelsimme arviota (21) viimeisessä epäyhtälössä. Palautetaan myös mieleen, että M_2 esittää tässä ylärajaa funktion f toisen derivaatan f'' itseisarvolle välillä $[a, b]$. Mikäli koko väli $[a, b]$ on jaettu pienempiin osaväleihin I_j , joiden pituus on Δx , niin yllä olevaa arviota voidaan soveltaa myös jokaiseen osaväliin I_j erikseen. Tämä tuottaa lopullisen virhearvion

$$\begin{aligned} \left| \int_a^b f(x) dx - \Delta x \left(\frac{f(a)}{2} + \sum_{j=2}^n f(x_j) + \frac{f(b)}{2} \right) \right| \\ \leq \sum_{j=1}^n \frac{M_2}{6} (\Delta x)^3 \end{aligned}$$

$$= \frac{M_2(b-a)}{6} (\Delta x)^2,$$

missä lopussa sievensimme $n\Delta x = b - a$. Puolisuunnikassäännössä tapahtuva virhe on siis yleisesti verrannollinen osavälien pituuden neliöön $(\Delta x)^2$. Tämän virheen voi myös ilmaista suhteessa laskettavien funktioiden arvojen määrään. Koska menetelmässä lasketaan funktion f arvoja yhteensä $n + 1$ kertaa ja $\Delta x = (b - a)/n$, niin virhe on verrannollinen myös laskettavien arvojen lukumäärän toiseen negatiiviseen potenssiin. Siis jos arvoja lasketaan kaksi kertaa enemmän, virhe pienenee noin neljäsosaan.

Osoittautuu, että myös suorakaidesäännön paras virhearvio on täysin samaa suuruusluokkaa. Tämä perustuu jälleen siihen ideaan, että suorakaidesäännön vakiofunktion voi korvata millä tahansa välin keskipisteeseen sovitetulla lineaarisella funktiolla. Sopivien kumoutumisten takia menetelmän aste siis oleellisesti kasvaa yhdellä. Jätämme kuitenkin yksityiskohdat harjoitustehtäväksi, sillä käymme seuraavaksi täysin saman ilmiön läpi Simpsonin säännössä.

Simpsonin säännön virhearvio

Simpsonin säännön virhearviointi on hieman mielenkiintoisempaa. Suora lasku soveltaen epäyhtälöä (21) kuten edellä antaisi virhearvion suuruusluokaksi $\approx (\Delta x)^3$, eli yhtä potenssia paremman kuin puolisuunnikassäännössä, mutta pystymme vielä parempaan. Perusidea on sama kuin suunnikassääntöä ja puolisuunnikassääntöä verrattaessa, sillä toisen asteen interpolaatiopolynomi tuleeikin tuottamaan samantarkkuisia arvioita kuin kolmannen asteen interpolaatiopolynomit.

Kuten ensimmäisen asteen tapauksessa, kolmannen asteen polynomeja integroitaessa tapahtuu hieman supistumista asteen parittomuuden takia. Ilmaisemme tämän ajatuksen nyt tarkasti.

Tutkitaan tilannetta koko välillä $[a, b]$, ja oletamme tässä laskujen yksinkertaistamisen vuoksi, että välin keskipiste on nollassa, eli että väli on muotoa $[-h, h]$, missä $h > 0$. Yleisessä tapauksessa tähän tilanteeseen voidaan aina palata muuttujanvaihdolla $y = x - \frac{a+b}{2}$.

Olkoon $p_2(x)$ Simpsonin säännön toisen asteen polynomi, joka on sovitettu pisteisiin $(-h, f(-h))$, $(0, f(0))$, ja $(h, f(h))$. Olkoon lisäksi $q(x)$ kolmannen asteen polynomi, joka kulkee näiden samojen pisteiden kautta. Vaikka polynomilla $q(x)$ on yksi ylimääräinen vapausaste, jota voi teoriassa hyödyntää esimerkiksi sovitamalla se yhden ylimääräisen pisteen kautta, ei tämä loppujen lopuksi vaikuta integraalin arvoon välin $[-h, h]$ ylitse. Tämä voidaan päätellä siitä, että erotuspolynomi $q(x) - p_2(x)$ on edelleen kolmannen asteen polynomi, jonka juuret ovat

$-h, 0$, ja h . Siis tekijöihin jaettuna

$$q(x) - p_2(x) = C(x-h)x(x+h) \quad \text{jollain vakiolla } C.$$

Polynomifunktio $C(x-h)x(x+h)$ on pariton, joten sen integraali symmetrisen välin yli on 0, ja voidaan laskea

$$\int_{-h}^h q(x) dx = \int_{-h}^h p_2(x) dx + \int_{-h}^h C(x-h)x(x+h) dx = \int_{-h}^h p_2(x) dx.$$

Siispä polynomien q ja p_2 integraalit välin $[-h, h]$ yli ovat samat.

Lopputuloks on, että Simpsonin säännössä tapahtuva virhe välillä $[a, b]$ voidaan ilmaista muodossa

$$\left| \int_a^b f(x) dx - \int_a^b p_2(x) dx \right| = \left| \int_a^b f(x) dx - \int_a^b q(x) dx \right|,$$

missä $q(x)$ on kolmannen asteen polynomi, joka saa samat arvot kuin f pisteissä a, b , ja $\frac{a+b}{2}$. Voidaan nyt sovittaa q yhden ylimääräisen pisteen kautta, esimerkiksi niin, että se saa saman arvon kuin f välin neljännesosassa $\frac{3a+b}{4}$. Nyt epäyhtälöä (21) soveltamalla saadaan

$$\left| \int_a^b f(x) - q(x) dx \right| \leq \frac{M_4(b-a)^5}{120}. \quad (25)$$

Tätä ideaa voidaan edelleen soveltaa myös välin $[a, b]$ osaväleihin, jolloin arvioksi saadaan suoraviivaisella laskulla

$$\left| \int_a^b f(x) dx - \frac{\Delta x}{6} \left[f(x_1) + 4 \sum_{j=1}^n f(m_j) + 2 \sum_{j=2}^n f(x_j) + f(x_{n+1}) \right] \right| \leq \frac{M_4(b-a)}{120} (\Delta x)^4.$$

Saimme siis kumoutumisten ansiosta parannettua approksimaation astetta yhdellä, ja tämä paransi myös virhetermin arvioissa osavälin pituuden potenssia yhdellä. Ilmaistaan virhe vielä laskettavien funktion arvojen lukumäärän suhteen. Koska laskettavia arvoja on yhteensä $2n+1$ kappaletta ja $\Delta x = (b-a)/n$, voidaan virheen yläraja esittää myös muodossa

$$\frac{M_4(b-a)}{120} (\Delta x)^4 = \frac{M_4(b-a)^5}{120} \frac{1}{n^4} \approx \frac{2M_4(b-a)^5}{15} \frac{1}{(2n+1)^4},$$

mistä nähdään, että virhe on verrannollinen laskettavien arvojen neljänteen negatiiviseen potenssiin. Siispä siinä missä suorakaide- ja puolisuunnikkasäännön virhe pieneni neljäsosaan kaksinkertaistamalla laskettavien arvojen määrän, niin Simpsonin säännössä virhe tipahtaa kuusitoista kertaa pienemmäksi. Simpsonin

sääntö on siis huomattavasti tehokkaampi, mutta vaatii toki lisäksi, että funktion f neljäs derivaatta käyttäytyy riittävän kesyesti välillä $[a, b]$.

Yleisesti ottaen Simpsonin säännölle voi todistaa vielä hivenen parempia virhearvioita (nimittäjän vakiota 120 voi parantaa entisestään), mutta nämä menevät melko teknisiksi laskuiksi. Tutkitaan sen sijaan edellä esiteltyjä menetelmiä vielä yhdestä uudesta näkökulmasta.

7.5 Gaussin integrointisääntö

Kaikki kolme edellä esitettyä numeerista integroimismenetelmää perustuivat lopulta integroitavan funktion f arvojen laskemiseen tietyissä välin $[a, b]$ pisteissä ja arvojen summaamiseen tietyillä kertoimilla. Esimerkiksi Simpsonin säännössä (24) pisteiksi valittiin välin päätepisteet ja keskipiste, ja kertoimet määräytyivät toisen asteen polynomin sovituksesta.

Yleisessä muodossa tällainen arviointimenetelmä voidaan kirjoittaa muodossa

$$\int_a^b f(x) dx \approx \sum_{k=1}^n w_k f(x_k), \quad (26)$$

missä pisteet x_k on valittu väliltä $[a, b]$, ja kertoimet w_k tulee myös valita sopivasti. Edellisissä menetelmissä valitsimme pisteet x_k tasavälein ja määritimme kertoimet w_k hyvän arvion tuottamiseksi, mutta entä jos myös pisteiden x_k sijanti voidaan valita vapaasti? Osoittautuu, että näin voidaan saavuttaa vielä parempia arvioita samalla pisteiden lukumäärällä kuin aiemmin. Tätä menetelmää kutsutaan **Gaussin säännöksi**.



Kuva 29. Carl Friedrich Gauss (1777–1855)

Haluamme yksinkertaistaa seuraavia laskuja, joten oletetaan, että välimme $[a, b]$ on origokeskeinen väli $[-1, 1]$. Tämä ei vähennä yleisyyttä, sillä jokainen väli $[a, b]$

saadaan välistä $[-1, 1]$ lineaarisella koordinaattimuunnoksella $y = \frac{b-a}{2}x + \frac{a+b}{2}$, missä $-1 \leq x \leq 1$ ja $a \leq y \leq b$. Tällä koordinaattimuunnoksella minkä tahansa funktion $f(x)$ määrätty integraali yli välin $[a, b]$ voidaan palauttaa integraaliksi yli välin $[-1, 1]$ muuttujanvaiholla

$$\int_a^b f(x) dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{b-a}{2}t + \frac{a+b}{2}\right) dt. \quad (27)$$

Integraalien muuttujanvaihtokaava käsitellään liitteessä B. Siis Gaussin sääntöä varten riittää tutkia pelkästään integraaleja välin $[-1, 1]$ yli.

Kysymys siis kuuluu: Jos meillä on n pistettä x_k käytettävissä, miten pisteet ja kertoimet w_k kannattaa valita niin, että arvio (26) on mahdollisimman tarkka? Yleisesti ottaen tämä toki riippuu funktiosta f , mutta pyritään ensin määrittämään pisteet ja kertoimet niin, että arvio (26) tepsii hyvin silloin, kun f on polynomi-funktio.

Osoittautuu, että valitsemalla pisteet x_k fiksusti voidaan kaavan (26) arvio muuttaa yhtäsuuruudeksi jopa polynomeille, jotka ovat selvästi suurempaa astetta kuin pisteiden lukumäärä n .

Lause: Gaussin sääntö. On mahdollista valita n pistettä x_k väliltä $[-1, 1]$, ja kertoimet w_k , joille kaava

$$\int_{-1}^1 p(x) dx = \sum_{k=1}^n w_k p(x_k) \quad (28)$$

on tarkka jokaiselle polynomille p , jonka aste on enintään $2n - 1$.

Tämä kaava on lisäksi yksikäsitteinen, eli pisteitä x_k ja kertoimia w_k ei voi valita kuin yhdellä tietyllä tavalla. Emme mene tässä Gaussin säännön todistuksen yksityiskohtiin, sillä ne ovat melko teknisiä, mutta lähestymme kaavaa esimerkkien kautta.

Gaussin säännön asteluvun ylärajalle $2n - 1$ on myös järkevä heuristinen perustelu miettimällä yhtälön (28) vapausasteita. Yleisellä polynomilla astetta $2n - 1$ on yhteensä $2n$ vapaasti säädettävää parametria, sen kertoimet $a_0, a_1, \dots, a_{2n-1}$. Sen sijaan yhtälön (28) oikealla puolella on myös $2n$ säädettävää muuttujaa, pisteet x_k ja kertoimet w_k . Tämä yhtäläisyys ei itsessään takaa ratkaisun olemassaoloa, mutta se on enne oikeaan suuntaan. Tutkitaan asiaa lähemmin yksinkertaisten tapausten kautta.

Aloitetaan tapauksesta $n = 1$, ja tällöin myös $2n - 1 = 1$. Halutaan siis valita w_1 ja x_1 siten, että yhtälön (28) vasemmasta puolesta tulee yhtä suurta kuin $w_1 p(x_1)$ kaikilla ensimmäisen ja nollannen asteen polynomeilla p . Riittää oikeastaan tarkistaa, että se pätee monomeille x ja 1 , sillä yleinen tapaus seuraa kertomalla

nämä vakioilla ja summaamalla yhteen. Asetetaan nämä yhtälöön (28), ja saadaan kaksi ehtoa

$$0 = \int_{-1}^1 x \, dx = w_1 p(x_1) = w_1 x_1 \quad \text{ja} \quad 2 = \int_{-1}^1 1 \, dx = w_1 p(x_1) = w_1 \cdot 1.$$

Näistä saadaan ratkaistua $w_1 = 2$ ja $x_1 = 0$. Siis kaava

$$\int_{-1}^1 p(x) \, dx = 2p(0)$$

pätee, kun polynomin p aste on enintään 1.

Tutkitaan nyt tapausta $n = 2$, jolloin $2n - 1 = 3$. Väitämme siis, että jopa kolmannen asteen polynomin integroimiseen riittää tietää sen arvo vain kahdessa tietyssä pisteessä. Tehdään tässä fiksu veikkaus, että $x_2 = -x_1$ ja $w_2 = w_1$, jonka idea paljastuu pian. Sijoitetaan kaavaan (28) taas monomifunktiot x^3 , x^2 , x , ja 1, ja saadaan

$$\begin{aligned} 0 &= \int_{-1}^1 x^3 \, dx = w_1 x_1^3 + w_1 (-x_1)^3 = 0 \\ \frac{2}{3} &= \int_{-1}^1 x^2 \, dx = w_1 x_1^2 + w_1 (-x_1)^2 = 2w_1 x_1^2 \\ 0 &= \int_{-1}^1 x \, dx = w_1 x_1 + w_1 (-x_1) = 0 \\ 2 &= \int_{-1}^1 1 \, dx = w_1 + w_1 = 2w_1. \end{aligned}$$

Ensimmäinen ja kolmas rivi siis toteutuivat automaattisesti, sillä aiempi arvauksemme parametrien x_2 ja w_2 valinnalle tuotti oikealla puolella saman kumoutumisen kuin parittoman asteen monomia integroitaessa välin $[-1, 1]$ yli. Mainittakoon, että sama ilmiö tapahtuu myös korkeamman asteen Gaussin säännöissä, ja pisteet x_k kannattaa aina valita symmetrisesti nollan ympäriltä, eli niin, että parittoman asteen termit kumoutuvat suoraan pois.

Jäljelle jäävistä kahdesta yhtälöstä voidaan ratkaista $w_1 = 1$ ja $x_1 = 1/\sqrt{3}$. Siis kaava

$$\int_{-1}^1 p(x) \, dx = p\left(\frac{1}{\sqrt{3}}\right) + p\left(-\frac{1}{\sqrt{3}}\right)$$

pätee kaikille kolmannen ja pienemmän asteen polynomeille p . Vastaavalla tavalla voidaan määrittää myös korkeamman asteen Gaussin säännön kaavoja. Näille kaavoille löytyy myös suurempia menetelmiä niin sanottujen Legendren polynomien avulla, joihin myös Gaussin säännön todistus perustuu.

n	$2n - 1$	Pisteet (x_i)	Painot (w_i)
1	1	0	2
2	3	$\pm \frac{1}{\sqrt{3}}$	1
3	5	0 $\pm \sqrt{\frac{3}{5}}$	$\frac{8}{9}$ $\frac{5}{9}$
4	7	$\pm \sqrt{\frac{3}{7} - \frac{2}{7}\sqrt{6/5}}$ $\pm \sqrt{\frac{3}{7} + \frac{2}{7}\sqrt{6/5}}$	$\frac{18+\sqrt{30}}{36}$ $\frac{18-\sqrt{30}}{36}$
5	9	0 $\pm \frac{1}{3}\sqrt{5-2\sqrt{10/7}}$ $\pm \frac{1}{3}\sqrt{5+2\sqrt{10/7}}$	$\frac{128}{225}$ $\frac{322+13\sqrt{70}}{900}$ $\frac{322-13\sqrt{70}}{900}$

Taulukko 1. Gaussin säännön pisteet ja kertoimet ensimmäiselle viidelle eri pisteiden lukumäärälle n . Kaava on tarkka polynomeille astetta $2n - 1$ tai vähemmän. Käytännön laskuissa käytetään luonnollisesti lukujen x_i ja w_i likiarvoja.

Gaussin säännön virhearvio

Koska Gaussin sääntö on tarkka kaikille polynomeille astetta $2n - 1$, tuottaa sen antama arvio myös hyviä approksimaatioita yleisemmille funktioille. Mikäli kertoimet w_k ja pisteet x_k ovat kuten Gaussin säännössä, pätee yleinen virhearvio

$$\left| \int_{-1}^1 f(x) dx - \sum_{k=1}^n w_k f(x_k) \right| \leq \frac{M_{2n} 2^{2n+1}}{(2n)!}.$$

Tämä voidaan jälleen todistaa suoraan arviosta (21), sillä jos Gaussin säännön arvio pätee tarkkana kaikille polynomeille enintään astetta $2n - 1$, niin se pätee myös tarkasti funktion f Lagrangen interpolaatiopolynomille valitsemissamme $2n$ pisteessä. Siis epäyhtälöä (21) voi soveltaa suoraan, ja saadaan ylläoleva virhearvio.

Gaussin säännön pisteitä ja kertoimia on hieman työläämpi laskea, kun n on suuri, mutta hyvän virhearvion ansiosta menetelmä on erittäin hyödyllinen kun sovelluksissa tarvitaan tarkkoja arvioita. Toinen tyyppillisesti sovelluksissa käytössä oleva integroimismenetelmä on johdannainen Simpsonin säännöstä. Niin sanotussa **dynaamisessa Simpsonin säännössä** osavälejä I_j ei valita välttämättä tasapituisesti, vaan niitä valitaan tiheämmin siellä missä integroimisvirhe on suurimmillaan. Tämä on laskentatehokkuuden ja virheen minimoimisen kannalta tehokas ja helposti tietokoneella toteutettava menetelmä.

Harjoitustehtäviä

Tehtävä 54. Laske Simpsonin säännön antama arvio funktion $f(x) = e^x$ integraalille välillä $[0, 1]$ ilman osavälejä, ja selvitä, kuinka iso virhe on todelliseen integraalin arvoon verrattuna.

Vihje s. 160 →

Toinen vihje s. 176 →

Tehtävä 55. Kun ympyrän säde on 1, sen ala on $\pi r^2 = \pi \cdot 1^2 = \pi$. Piin likiarvo voidaan siis selvittää laskemalla yksikköpuoliympyrän pinta-ala integroimalla ja kertomalla tulos kahdella:

$$\pi = 2 \int_{-1}^1 \sqrt{1-x^2} dx.$$

Laske luvulle π likiarvo approksimoimalla integraalia

- 5 pisteen eli 4 jakovälin puolisuunnikassäännöllä,
- 101 pisteen eli 100 jakovälin puolisuunnikassäännöllä,
- 101 pisteen Simpsonin säännöllä, ja
- 5 pisteen Gaussin säännöllä.

Huomautettakoon, että tämä ei ole kovin tehokas tapa approksimoida lukua π . Ympyrän pystysuora reuna kohdissa $x = \pm 1$ tuottaa integrointimenetelmille ongelmia.

Vihje s. 161 →

Toinen vihje s. 176 →

Tehtävä 56. Tietokoneet laskevat logaritmillemme likiarvoja esimerkiksi Taylorin polynomien avulla. Tämä onnistuu myös numeerisella integroinnilla. Pätee nimittäin

$$\int_1^a \frac{1}{x} dx = \int_1^a \ln x = \ln a - \ln 1 = \ln a.$$

- Laske luvulle $\ln 2$ likiarvo laskemalla edellä mainittu integraali neljän osavälin puolisuunnikassäännöllä. Vertaa tarkempaan likiarvoon. Miksi tällä menetelmällä saatava tulos on varmasti hieman liian suurempi kuin $\ln 2$?
- Hyödynnä sivulla 97 esiteltyä puolisuunnikassäännön virhekaavaa ja arvioi sen perusteella, kuinka moneen osaväliin väli $[1, 2]$ tulisi jakaa, jotta integroimalla saatu luvun $\ln 2$ likiarvo poikkeaisi varmasti oikeasta arvosta vähemmän kuin 10^{-6} .

Vihje s. 161 →

Toinen vihje s. 176 →

Tehtävä 57. Laske tarkka kaava Simpsonin säännön virheelle välillä $[-1, 1]$, kun integroitava funktio $f(x)$ on neljännen asteen polynomi. Vertaa tätä vielä virhearvion (25).

Vihje s. 161 →

Toinen vihje s. 177 →

Tehtävä 58. Laske tarkka kaava puolisuunnikassäännön antamalle likiarvolle, kun integroidaan funktiota $f(x) = x^2$ välillä $[0, 1]$, ja osavälejä on n kappaletta. Voit pitää tunnettuna kaavan

$$1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Tarkista, että löytämäsi arvion lausekkeen raja-arvo, kun $n \rightarrow \infty$, antaa todellisen integraalin arvon $\frac{1}{3} = \int_0^1 x^2 dx$.

Vihje s. 161 →

Toinen vihje s. 177 →

Tehtävä 59. Laske kahden pisteen Gaussin säännöllä integraalille $\int_2^3 x \ln(x) dx$ likiarvo. Vertaa arviota integraalin todelliseen arvoon.

Vihje s. 161 →

Toinen vihje s. 177 →

Tehtävä 60. Johda Gaussin sääntö kolmelle pisteelle, eli määritä kolme pistettä x_k ja vastaavat kertoimet w_k niin, että kaava (28) pätee jokaiselle viidennen asteen polynomille. Lopullinen vastaus löytyy jo luvussa esiintyneestä taulukosta, mutta johda se itse.

Vihje s. 161 →

Toinen vihje s. 177 →

Tehtävä 61. Aina numeerisen integroinnin tarkkuutta ei voida tai haluta arvioida virhekaavalla. Virheen kokoluokkaa voi kuitenkin arvioida seuraavalla tempulla:

Kuvitellaan, että integraalin tarkka arvo on $I = \int_a^b f(x) dx$, ja sille on laskettu puolisuunnikassäännöllä likiarvo I_n käyttäen n jakoväliä. Pätee siis

$$I = I_n + \varepsilon_n,$$

missä ε_n on numeerisen integroinnin virhe. Verrataan tulosta kaksinkertaisella määrällä jakovälejä laskettuun likiarvoon I_{2n} , jonka virhe on ε_{2n} . Koska puolisuunnikassäännössä virheen yläraja on verrannollinen jakovälin pituuden neliöön, virheen ε_n yläraja on nelinkertainen virheen ε_{2n} ylärajaan nähden. Tämä antaa motivaation approksimoida $\varepsilon_n \approx 4\varepsilon_{2n}$. Tämä approksimaatio ei aina ole hyvä, sillä nelinkertainen virheen yläraja ei automaattisesti tarkoita nelinkertaista virhettä. Joka tapauksessa likimääräisestä yhtälöparista voidaan ratkaista

$$\begin{cases} I \approx I_n + 4\varepsilon_{2n} \\ I = I_{2n} + \varepsilon_{2n} \end{cases} \Rightarrow \varepsilon_{2n} \approx \frac{I_{2n} - I_n}{3}.$$

Arvioi tällä menetelmällä virhettä, joka syntyy, kun integraalille $\int_0^\pi \sin x \, dx$ lasketaan likiarvo neljän jakovälin puolisuunnikassäännöllä.

Vihje s. 161 →

Toinen vihje s. 177 →

Tehtävä 62. Tehtävän 61 tehtävänannossa esiteltiin arvio $\varepsilon_{2n} \approx (I_{2n} - I_n)/3$, missä I_n ja I_{2n} ovat puolisuunnikassäännöllä jakovälien määrillä n ja $2n$ laskettuja likiarvoja integraalille $I = \int_a^b f(x) \, dx$, ja ε_{2n} on likiarvon I_{2n} virhe. Nämä yhdistämällä saadaan arvio

$$I = I_{2n} + \varepsilon_{2n} \approx I_{2n} + \frac{I_{2n} - I_n}{3}.$$

Osoita, että tämä arvio on itse asiassa sama kuin Simpsonin säännön kaava.

Vihje s. 162 →

Toinen vihje s. 178 →

8 Differentiaaliyhtälöt

Differentiaaliyhtälö on matemaattinen yhtälö, jossa esiintyy tuntematon funktio sekä sen mahdollisia derivaattoja. Tyypillinen esimerkki voisi olla vaikkapa yhtälö

$$y'(x) = x \cdot y(x),$$

josta halutaan ratkaista funktio y . Kyse on siis tilanteesta, jossa pyrimme selvittämään mitkä funktiot y toteuttavat annetun vuorovaikutuksen funktion y ja sen derivaattojen $y', y'', \dots$ välillä. Huomaa, että tässä y' tarkoittaa funktion $y(x)$ derivaattaa muuttujan x suhteen, eli $y' = \frac{d}{dx} y$.

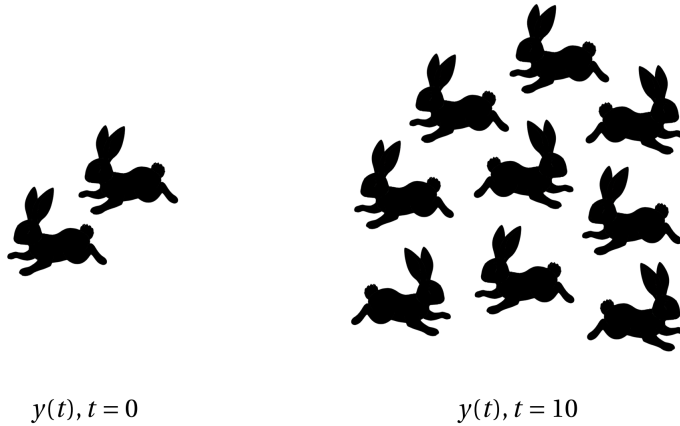
Esimerkkiyhtälömme eräs ratkaisu olisi vakiofunktio $y(x) = 0$, ja toinen ratkaisu $y(x) = e^{\frac{1}{2}x^2}$, sillä sen derivaataksi saadaan $y'(x) = \frac{d}{dx} e^{\frac{1}{2}x^2} = x \cdot e^{\frac{1}{2}x^2} = x \cdot y(x)$.

Osoittautuu, että differentiaaliyhtälöt ovat aivan loistava työkalu mallintamaan monenlaisia todellisen maailman ilmiöitä kuten planeettojen liikkeitä, lämmön jakautumista, molekyylien käyttäytymistä tai populaatioiden muutoksia. Tästä syystä differentiaaliyhtälöt ovat keskeisessä asemassa monissa luonnontieteiden ja tekniikan osa-alueissa.

Lähdetään tutkimaan konkreettisen esimerkin kautta, miten differentiaaliyhtälö voidaan muodostaa ratkaisemaan käytännön mallinnusongelma ja esitetään yksi ratkaisumenetelmä.

Populaatioesimerkki 1. Metsässä asuu jäniksiä, joiden lukumäärää ajanhetkellä $t \geq 0$ kuvaa tuntematon funktio $y(t)$. Toki kun kyse on yksilöiden lukumäärästä, olisi luonnollisinta olettaa, että y saa arvoja vain positiivisten kokonaislukujen joukossa. Mutta hyväksytään, että kyseessä on pelkkä approksimaatio ja sallitaan, että y saa myös ei-kokonaislukuarvoja. Populaatiota eri ajanhetkillä kuvataan siis funktiona $y: [0, \infty[\rightarrow [0, \infty[$. Esimerkkimme ei pyri muutenkaan kuvaamaan realistisesti minkään todellisen populaation kehitystä vaan antamaan karkean ylätasoin mallin.

Jänisten populaatio kasvaa, kun uusia yksilöitä syntyy. Syntyvyys ei kuitenkaan ole vakio, vaan oletetaan sen riippuvan jänispopulaation koosta $y(t)$. Tämä käy järkeen, sillä jos ravintoa riittää, suurempi määrä jäniksiä voi tuottaa suuremman määrän poikasia. On siis luontevaa olettaa, että millä tahansa ajanhetkellä t jänisten syntyvyys on suoraan verrannollinen populaation kokoon $y(t)$, eli löytyy jokin vakio $B > 0$ jolla syntyvyys hetkellä t on suuruudeltaan $By(t)$.



Vastaavasti jäniksiä kuolee tasaisesti luonnollisiin syihin. Pidetään malli yksinkertaisena, joten ei huolehdi vielä siitä, miten isompi jänispopulaatio todellisuudessa johtaa resurssien loppumiseen, petoeläinten lisääntymiseen ja sairauksien leviämiseen.

Oletetaan sen sijaan, että kuolevuus on myös suoraan verrannollinen populaation lukumäärään $y(t)$, eli löytyy jokin vakio $D > 0$ jolla kuolevuus hetkellä t on suuruudeltaan $Dy(t)$.

Jänispopulaation muutosnopeutta kuvastaa populaation lukumäärän $y(t)$ derivaatta muuttujan t suhteen, jota merkitään taas $y'(t)$. Toisaalta muutosnopeus on myös sama asia kuin syntyvyyden ja kuolevuuden erotus, joten voimme muodostaa differentiaaliyhtälön

$$y'(t) = By(t) - Dy(t) = (B - D)y(t).$$

Jos merkitään $k = B - D$, niin yhtälö sievenee muotoon

$$y'(t) = ky(t). \quad (29)$$

Pyritään nyt ratkaisemaan minkälaisilla funktioilla y tämä yhtälö toteutuu. Arvaamalla voi jo huomata, että eksponenttifunktio $y(t) = e^{kt}$ on yksi potentiaalinen ratkaisu, mutta pyritään etsimään kaikki ratkaisut. Populaation määrä $y(t)$ on oletettavasti positiivinen, joten voidaan jakaa sillä yhtälö puolittain ja saattaa se muotoon

$$\frac{y'(t)}{y(t)} = k.$$

Nyt on kriittistä huomata, että vasemman puolen lauseke $y'(t)/y(t)$ on erään yhdistetyn funktion derivaatta. Nimittäin jos funktiota $\ln(y(t))$ derivoi, niin yhdistetyn funktion derivointikaavalla saadaan vastaukseksi $\frac{d}{dt}(\ln y(t)) = 1/y(t) \cdot y'(t) = y'(t)/y(t)$.

Vastaavasti oikean puolen lauseke, vakio k , on funktion kt derivaatafunkti muuttujan t suhteen. Siispä yhtälö voidaan myös kirjoittaa muodossa

$$\frac{d}{dt}(\ln y(t)) = \frac{d}{dt}(kt) \quad \text{tai vastaavasti} \quad \frac{d}{dt}(\ln y(t) - kt) = 0.$$

Tästä muodosta differentiaaliyhtälön ratkaisu on helppo selvittää, sillä ainoa funktio, jonka derivaatta on nolla, on vakiofunktio. Siis löytyy jokin reaalinen vakio C , jolle pätee

$$\ln y(t) - kt = C,$$

mistä voidaan ratkaista y siirtämällä kt oikealle puolelle ja ottamalla eksponenttifunktio puolittain (muistaen, että $e^{\ln x} = x$ kun $x > 0$). Tämä antaa ratkaisuksi

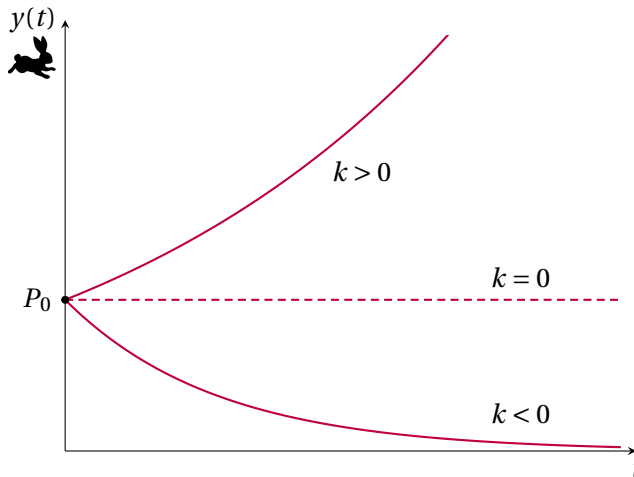
$$y(t) = e^{kt+C}.$$

Ratkaisua voi halutessaan vielä sieventää. Jos merkitään $P_0 = e^C$, niin eksponenttifunktion laskukaavoilla saadaan

$$y(t) = e^{kt+C} = e^C e^{kt} = P_0 e^{kt}.$$

Tässä P_0 itse asiassa kuvastaa jänispopulaation kokoa alkuhetkellä $t = 0$, sillä $y(0) = P_0 e^0 = P_0$. Huomataan myös, että tämä alkupopulaatio on positiivinen, sillä $P_0 = e^C$ ja eksponenttifunktio saa vain positiivisia arvoja.

Ratkaisusta voimme nyt päätellä yhtä ja toista jänisten populaatiosta. Jos syntyvyyden ja kuolleisuuden erotus on positiivinen, niin $k = B - D > 0$ ja tällöin jänispopulaation koko $y(t) = P_0 e^{kt}$ kasvaa eksponentiaalista vauhtia ylöspäin. Jos $k = 0$, eli syntyvyys ja kuolevuus ovat täsmälleen samat, niin luonnollisesti $y(t) = P_0 e^0 = P_0$ pysyy vakiona. Jos taas $k < 0$, niin jänisten lukumäärä lähestyy nopeasti nollaa.



Tässä yksinkertaisessa esimerkissä tutustuimme yhteen keskeiseen tapaan ratkaista differentiaaliyhtälö: muokataan yhtälöä, kunnes molemmilta puolilta löytyy jonkin funktion derivaattafunktio. Tutkitaan seuraavaksi, missä yleisyydessä tätä tekniikkaa voidaan soveltaa.

8.1 Separoituvat differentiaaliyhtälöt

Tehdään ensin pieni sopimus notaation suhteen. Differentiaaliyhtälöitä kirjoittaessa on tyypillistä jättää muuttuja pois kun kirjoitetaan ratkaistava funktio. Siis esimerkiksi yhtälö

$$y'(x) = x \cdot y(x) \quad \text{voidaan kirjoittaa} \quad y' = xy.$$

Kontekstista kuitenkin ymmärretään, että yhtälöstä $y' = xy$ ratkaistava funktio y riippuu muuttujasta x .

Määritelmä. Differentiaaliyhtälöä sanotaan **separoituvaksi differentiaaliyhtälöksi**, jos se on muotoa

$$y' = f(x)g(y), \tag{30}$$

missä f ja g ovat annettuja funktioita, ja y on ratkaistava tuntematon muuttujan x funktio. Nimensä mukaisesti separoituvassa differentiaaliyhtälössä funktiosta y ja muuttujasta x riippuvat tekijät voidaan erottaa eri puolille yhtälöä yhdellä jakolaskulla.

Esimerkki. Kaikki seuraavat yhtälöt ovat separoituvia.

$$\begin{array}{ll} y' = e^x y, & \text{missä } f(x) = e^x, g(y) = y \\ y' = \frac{y^2}{x}, & \text{missä } f(x) = \frac{1}{x}, g(y) = y^2 \\ y' = \frac{e^x}{y^2 - y + 1}, & \text{missä } f(x) = e^x, g(y) = (y^2 - y + 1)^{-1} \end{array}$$

Separoituvan differentiaaliyhtälön ratkaiseminen

Separoituvan yhtälön $y' = f(x)g(y)$ ratkaisemista varten tarvitsemme funktioille f ja $1/g$ integraalifunktiot F ja G . Oletetaan siis, että löytyy funktiot F ja G joille pätee

$$\frac{d}{dx} F(x) = f(x) \quad \text{ja} \quad \frac{d}{dy} G(y) = \frac{1}{g(y)}.$$

Huom! Yllä olevassa kaavassa y merkitsee funktioiden g ja G muuttujaa, ei vielä tuntematonta funktiota $y(x)$.

Ratkaistaan nyt separoituva yhtälö $y' = f(x)g(y)$. Tutkitaan ensin tapausta missä $g(y) \neq 0$, jolloin lausekkeella $g(y)$ voidaan jakaa ja saadaan yhtälö muotoon

$$\frac{y'}{g(y)} = f(x). \quad (31)$$

Jälleen yhdistetyn funktion derivointikaavalla nähdään, että

$$\frac{d}{dx}G(y(x)) = G'(y) \cdot y' = \frac{1}{g(y)} \cdot y',$$

joten integroimalla yhtälö (31) puolittain muuttujan x suhteen saadaan

$$\int \frac{1}{g(y)} \cdot y'(x) dx = \int f(x) dx$$

$$G(y) + C_1 = F(x) + C_2,$$

ja merkitsemällä $C = C_2 - C_1$ saamme separoituvan differentiaaliyhtälön ratkaisukaavan

$$G(y) = F(x) + C, \quad (32)$$

missä C on vapaasti valittava reaalinen vakio. Tällaisesta yhtälöstä on tyypillisesti helppo selvittää mahdolliset ratkaisut funktiolle y . Esimerkiksi jos funktiolla G on käänteisfunktio G^{-1} , niin ratkaisu on muotoa

$$y(x) = G^{-1}(F(x) + C).$$

Aiemmassa vaiheessa oletimme, että $g(y) \neq 0$. Tutkitaan vielä, mitä tapahtuu jos $g(y) = 0$. Tällöin funktiolla g täytyy olla nollakohta jossakin pisteessä y_0 . Huomataan, että tällöin vakiofunktio $y(x) \equiv y_0$ on itse asiassa alkuperäisen yhtälön (30) eräs ratkaisu. Merkinnällä $\equiv$ korostetaan, että yhtäsuuruus on voimassa kaikilla muuttujan x arvoilla.

Kun yhtälön määrittävät funktiot f ja g ovat riittävän säännöllisiä (tarkka käsittely sivulla 113), on mahdollista osoittaa, että mikään muu kuin vakioratkaisu ei saa funktion g nollakohtia arvokseen. Tällöin edellä käyty tapaustarkastelu kattaa kaikki mahdollisuudet, ja separoituvan yhtälömme ratkaisut löytyvät siis joko

1. yhtälöstä $G(y) = F(x) + C$, missä C on vapaavalintainen vakio, tai
2. vakioratkaisuna $y(x) \equiv y_0$, missä y_0 on funktion g nollakohta.

Esimerkki 1. Ratkaistaan separoituva differentiaaliyhtälö

$$y' = e^y.$$

Tässä yhtälössä $g(y) = e^y \neq 0$, joten voidaan jakaa tekijällä e^y ja saadaan separoitu muoto

$$e^{-y} y' = 1.$$

Integroidaan tämä muuttujan x suhteen, ja saadaan

$$\int e^{-y(x)} \cdot y'(x) dx = \int 1 dx$$

$$e^{-y(x)} = x + C,$$

missä olemme taas yhdistäneet integrointivakiot yhdeksi vakioksi C . Tästä saadaan yhtälön ratkaisuksi $y(x) = -\ln(x + C)$, missä $C \in \mathbb{R}$. Eräs yhtälön ratkaisu on siis esimerkiksi $y(x) = -\ln(x + 1)$.

Esimerkki 2. Ratkaistaan separoituva differentiaaliyhtälö

$$y' = \frac{y-1}{x}. \quad (33)$$

Huom! Ensin pieni huomautus yhtälön määrittelyjoukosta. Tämä yhtälö ei ole määritelty pisteessä $x = 0$, joten on luontevaa ratkaista yhtälö erikseen tapauksissa $x > 0$ ja $x < 0$. Tällaisessa tilanteessa, jossa yhtälön määrittelyjoukko jakautuu erillisiin palasiin, on riittävää ratkaista yhtälö joka palassa erikseen. Tarkemmin sanottuna, jos y_1 on yksi yhtälön (33) ratkaisu palassa $x \in]0, \infty[$ ja y_2 on jokin ratkaisu palassa $x \in]-\infty, 0[$, niin yleinen ratkaisu voidaan muodostaa kaavasta

$$y(x) = \begin{cases} y_1(x), & \text{kun } x > 0, \\ y_2(x), & \text{kun } x < 0. \end{cases}$$

Keskitytään tässä esimerkissä tapaukseen $x > 0$. Ensimmäisenä huomataan, että kun $y = 1$, yhtälön oikea puoli on nolla. Vakioratkaisu $y(x) \equiv 1$ on siis yksi yhtälön ratkaisu.

Kun $y \neq 1$, voidaan jakaa tekijällä $y - 1$ ja saadaan separoitu muoto, joka voidaan integroida muuttujan x suhteen:

$$\frac{y'}{y-1} = \frac{1}{x}$$

$$\int \frac{y'}{y-1} dx = \int \frac{1}{x} dx$$

$$\ln|y-1| = \ln x + C_0.$$

Huom! Tässä muuttujan x kanssa ei tarvita itseisarvoa, koska $x > 0$, mutta lauseke $y - 1$ vaatii itseisarvot ympärilleen, koska emme tiedä, onko $y - 1$ positiivinen vai negatiivinen.

Nyt voidaan ratkaista

$$|y - 1| = e^{\ln x + C_0} = x \cdot e^{C_0}.$$

Sievennetään merkitsemällä $C_1 = e^{C_0} > 0$, jolloin $|y - 1| = C_1 x$. Tästä saadaan kahdenlaisia ratkaisuja:

Tapaus 1: $y - 1 > 0$. Tällöin $y - 1 = C_1 x$, eli $y = 1 + C_1 x$ jollakin vakiolla $C_1 > 0$.

Tapaus 2: $y - 1 < 0$. Tällöin $-(y - 1) = C_1 x$, eli $y = 1 - C_1 x$ jollakin vakiolla $C_1 > 0$.

Yleinen ratkaisu voidaan nyt kirjoittaa muodossa $y = 1 + Cx$, missä C voi olla mikä tahansa reaali-luku. Tapaukset $C > 0$, ja $C < 0$ kattavat edellä käsitellyt tapaukset 1 ja 2, ja tapaus $C = 0$ vastaa vakioratkaisua $y(x) \equiv 1$.

Separoituvat differentiaaliyhtälöt ovat esimerkkejä **ensimmäisen asteen differentiaaliyhtälöistä**, eli niissä esiintyy ainoastaan tuntemattoman funktion y ensimmäisen asteen derivaatta y' eikä korkeamman asteen derivaattoja. Tutkitaan seuraavaksi, miten ratkaistaan separoituvan yhtälön jokin tietty ratkaisu.

8.2 Alkuarvo-ongelmat

Kuten aiemmissa esimerkeissä havaittiin, differentiaaliyhtälöitä ratkaistaessa vastaukseen jää usein vapaasti valittava vakioparametri, kuten kerroin P_0 jänispopulaatio-esimerkin ratkaisussa $y(t) = P_0 e^{kt}$. Yhden vapaan parametrin ilmestyminen on tyypillistä ensimmäisen asteen differentiaaliyhtälöitä ratkaistaessa. Tämä vapaa parametri mahdollistaa yleensä ratkaisufunktion $y(x)$ arvon määrittämisen vapaasti jossakin tietyssä pisteessä $x = x_0$. Tutkitaan esimerkin kautta, miten tämä toimii.

Esimerkki. Ratkaistaan alkuarvo-ongelma

$$\begin{cases} y' = y^2 \\ y(0) = 1. \end{cases}$$

Yhtälöllä $y' = y^2$ on vakioratkaisu $y(x) \equiv 0$, mutta se ei toteuta alkuarvoa $y(0) = 1$. Yhtälö on separoituva, missä $f(x) = 1$ ja $g(y) = y^2$. Integroimalla saadaan

$$\int \frac{y'}{y^2} dx = \int 1 dx - 1/y = x + C,$$

mistä edelleen voidaan ratkaista $y(x) = -1/(x + C)$. Jotta alkuarvo $y(0) = 1$ toteutuisi, sijoitetaan $x = 0$ ja saadaan $1 = -1/C$, eli $C = -1$. Siis alkuarvo-ongelman ainoa ratkaisu on funktio

$$y(x) = \frac{-1}{x - 1} = \frac{1}{1 - x}.$$

Huom! On myös mahdollista, että alkuarvo-ongelmalla ei ole ratkaisua, tai että se ei ole yksikäsitteinen. Kahdessa seuraavassa esimerkissä käy juuri niin.

Esimerkki. Alkuarvo-ongelmalla

$$\begin{cases} yy' = 1 \\ y(5) = 0 \end{cases},$$

ei ole ratkaisua, sillä jos y saa arvon 0 jossakin pisteessä, niin yhtälö väittää $0 = 1$. Sen sijaan alkuarvo-ongelmalla

$$\begin{cases} y' = \sqrt[3]{y} \\ y(0) = 0 \end{cases},$$

on useita eri ratkaisuja, esimerkiksi $y(x) \equiv 0$ ja $y(x) = x\sqrt{\frac{2}{3}|x|}$.

Edellisessä esimerkissä rikkoontui kaksi differentiaaliyhtälöiden toivottua perusominaisuutta: ratkaisun olemassaolo ja sen yksikäsitteisyys. Näin käy onneksi kuitenkin vain sellaisissa tapauksissa, jossa yhtälössä esiintyy tiettyä epäsäännöllisyyttä. Esimerkin yhtälöissä

$$y' = 1/y \quad \text{ja} \quad y' = \sqrt[3]{y}$$

molemmissa oikean puolen lauseke ei joko ollut jatkuva tai ei ollut derivoituva kohdan $y = 0$ läheisyydessä. Esitämme seuraavaksi tuloksen, joka sulkee sopivilla oletuksilla nämä epämukavat tapaukset pois.

Olemassaolo ja yksikäsitteisyys

On yleensä hyvin toivottua, että alkuarvo-ongelmalla olisi olemassa yksi ja vain yksi ratkaisu, varsinkin jos mallilla yritetään ennustaa jonkin oikean maailman ilmiön käyttäytymistä.

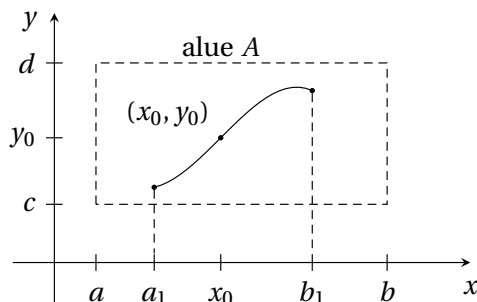
Yksi differentiaaliyhtälöiden matemaattisen teorian peruskysymyksistä onkin ratkaisujen olemassaolo ja yksikäsitteisyys annetuilla alkuarvoilla. Monimutkaisilla yhtälöillä tämän asian selvittäminen johtaa usein myös hankaliin tutkimuskysymyksiin. Tutustumme nyt erääseen perustulokseen, joka vastaa tähän kysymykseen laajassa luokassa ensimmäisen asteen differentiaaliyhtälöitä.

Olemassaololause (Cauchy ja Peano). Tutkitaan ensimmäisen asteen alkuarvo-ongelmaa, joka on muotoa

$$\begin{cases} y' = h(x, y) \\ y(x_0) = y_0 \end{cases}, \quad (34)$$

missä $h(x, y)$ on annettu muuttujien x ja y funktio, ja $(x_0, y_0) \in \mathbb{R}^2$ on annettu alkuarvopiste. Meidän täytyy myös kiinnittää, millä x :n ja y :n arvoilla funktio h

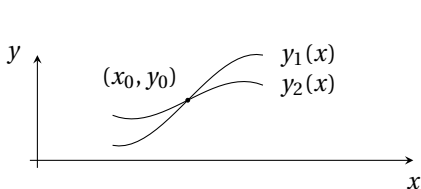
on määritelty. Oletetaan, että on olemassa reaaliluvut $a < b$ ja $c < d$ siten, että h on määritelty x :n arvoilla välillä $[a, b]$ ja y :n arvoilla välillä $[c, d]$. Siispä h :n määrittelyalue on suorakaide $A = [a, b] \times [c, d]$. On myös luontevaa olettaa, että alkuarvopiste (x_0, y_0) kuuluu aidosti tähän määrittelyalueeseen, eli että $a < x_0 < b$ ja $c < y_0 < d$.



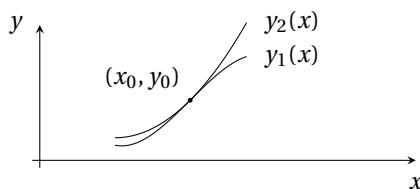
Olemassaolo. Cauchyn ja Peanon olemassaololause sanoo nyt seuraavaa: Jos kahden muuttujan funktio $h(x, y)$ on jatkuva alueessa A , niin alkuarvo-ongelmalla (34) on ainakin yksi ratkaisu jollakin osavälillä $a_1 \leq x \leq b_1$, missä

$$a_0 \leq a_1 < x_0 < b_1 \leq b_0.$$

Yksikäsitteisyys. Koska yhtälö $y' = h(x, y)$ määrää yksikäsitteisen derivaatan arvon jokaisessa tason pisteessä (x, y) , alkuarvo-ongelmalla ei voi olla ratkaisuja, jotka leikkaavat toisensa jossakin nollaa suuremmassa kulmassa. Tämä ei kuitenkaan sulje pois ratkaisuja, jotka sivuavat toisiaan. Voidaan kuitenkin osoittaa, että jos funktion $h(x, y)$ osittaisderivaatta muuttujan y suhteen, eli $\frac{\partial h}{\partial y}$, on jatkuva alueessa A , edellä mainittu ratkaisu on yksikäsitteinen.



Kuva 30. Ei näin! Alkuarvo-ongelman $y' = h(x, y)$, $y(x_0) = y_0$ ratkaisulla y_1 ja y_2 ei voi olla kahta eri derivaattaa samassa pisteessä.



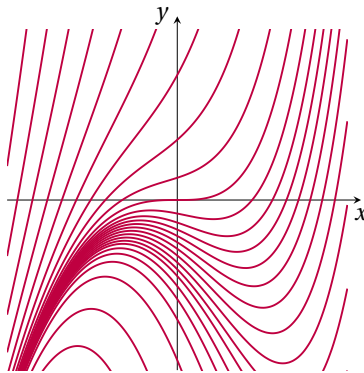
Kuva 31. Eikä edes näin, mikäli $\frac{\partial h}{\partial y}$ on jatkuva, sillä silloin ratkaisuja on vain yksi.

Tutkitaan vielä uudestaan, mitä lausetta edeltävän esimerkin alkuarvo-ongelmissa meni pieleen. Ensimmäisessä esimerkissä differentiaaliyhtälö $yy' = 1$ saadaan

haluttuun muotoon (34) valitsemalla $h(x, y) = 1/y$, mutta tässä tapauksessa alkuarvoa $y(0) = 0$ vastaava alkuarvopiste $(0, 0)$ ei kuulu funktion h määrittelyjoukkoon, ja funktiota h ei voida määritellä millään tavalla jatkuvaksi tässä pisteessä.

Toisen esimerkin yhtälössä $y' = h(x, y) = \sqrt[3]{y}$ funktio h on jatkuva, mutta osittaisderivaatta $\frac{\partial}{\partial y} h$ ei ole edes määritelty origossa. Yksikäsitteisyysehto ei siis täyty, joten alkuarvo-ongelman $y(0) = 0$ ratkaisuja voi olla – ja onkin – useampi. Vastaavaa esimerkkiä käsitellään tehtävässä 70.

Joka tapauksessa varsin usein sekä $h(x, y)$ että $\frac{\partial h}{\partial y}$ ovat jatkuvia, jolloin kutakin alkuarvoa vastaa varmasti yksikäsitteinen ratkaisu. Tällöin yhtälön $y' = h(x, y)$ eri ratkaisut eivät leikkaa toisiaan. Näin on esimerkiksi yhtälöllä $y' = x^2 + y$, jonka eri ratkaisuja on hahmoteltu seuraavaan kuvaajaan.



Differentiaaliyhtälön $y' = x^2 + y$ ratkaisuja

Mainittakoon vielä, että ratkaisun yksikäsitteisyyttä koskevasta lauseesta on olemassa myös vahvempia versioita. Yksikäsitteisyyteen riittää esimerkiksi se, että h on niin sanotusti Lipschitz-jatkuva. Tämä tunnetaan Picardin-Lindelöfin lauseena, jolle ovat antaneet nimensä ranskalainen Émile Picard (1856–1941) ja suomalainen Ernst Lindelöf (1870–1946).



Ernst Lindelöf vastavalmistuneena filosofian tohtorina vuonna 1894. [18]

8.3 Suuntakenttä

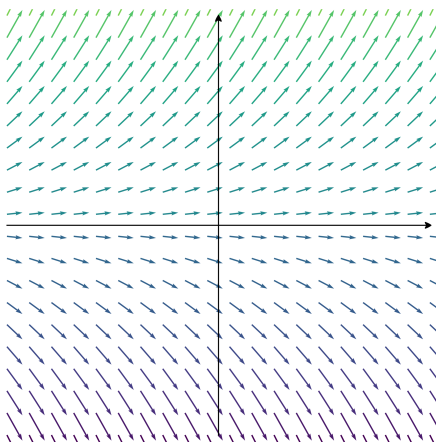
Koska differentiaaliyhtälöiden yleinen ratkaiseminen on usein haastavaa, on kehitetty monia vaihtoehtoisia tapoja ymmärtää ratkaisufunktioiden käyttäytymistä. Yksi tällainen tapa on **suuntakenttä**, jonka avulla voidaan hahmotella ratkaisun kuvaajan käyttäytymistä visuaalisesti.

Tutkitaan edelleen ensimmäisen asteen differentiaaliyhtälöä $y' = h(x, y)$. Edellä esitetty olemassaololause takaa, että mikäli funktio h käyttäytyy riittävän hyvin, on yhtälöllä olemassa jokaista tason pistettä (x_0, y_0) kohti ratkaisu, jonka kuvaaja kulkee tämän pisteen kautta (siis ratkaisu, jolla alkuarvo $y(x_0) = y_0$ toteutuu).

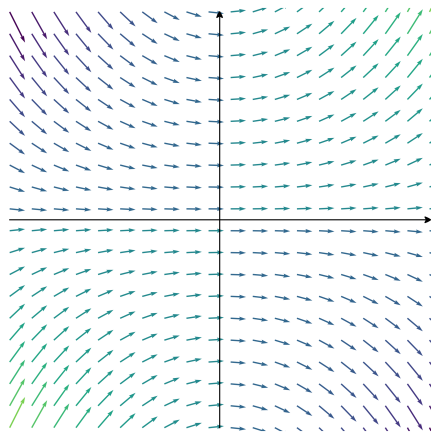
Ensimmäisen asteen differentiaaliyhtälö on siitä näppärä, että se paljastaa meille myös kyseisen ratkaisufunktion kuvaajan tangenttisuoran pisteessä (x_0, y_0) . Tämä tangentti kulkee nimittäin vektorin $(x_0, y'(x_0))$ suuntaisesti, ja tangenttivektorille saadaan yhtälöstä kaava

$$(x_0, y'(x_0)) = (x_0, h(x_0, y_0)).$$

Tästä kaavasta tangenttivektori voidaan laskea joka pisteessä (x_0, y_0) . Vektori ei

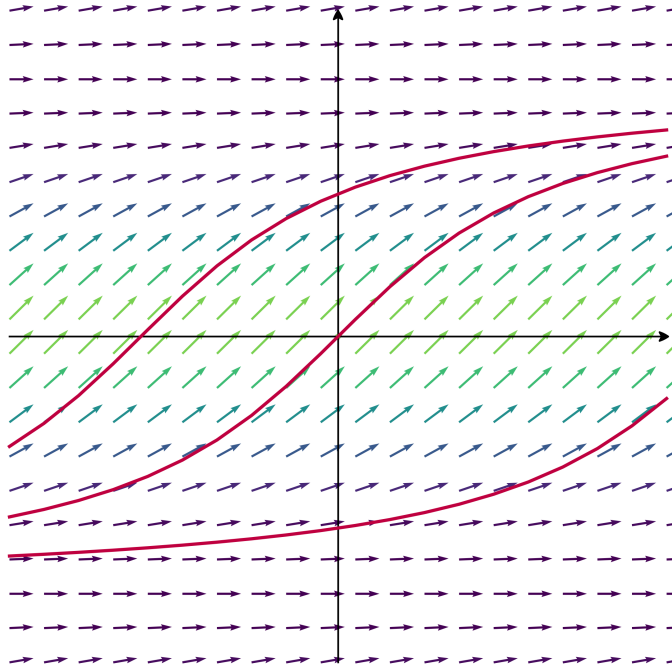


Kuva 32. Suuntakenttä yhtälölle $y' = y$, jonka ratkaisut ovat eksponenttifunktioita.



Kuva 33. Suuntakenttä yhtälölle $2y' = xy$, jonka ratkaisut ovat paraabeleita.

paljasta pelkästään tangentin suunnan, vaan myös ratkaisufunktion kasvunopeuden pisteessä x_0 , sillä vektorin y' -komponentti $y'(x_0)$ kuvaa tätä kasvunopeutta. Siis pidempi vektori tarkoittaa nopeampaa kasvua tai vähenemistä. Yhtälön suuntakenttä hahmotetaan piirtämällä tätä vektoria kuvaavia nuolia riittävän tiheään tason pistejoukkoon. Suuntakenttää katsomalla voi silmämääräisesti hahmottaa, miten differentiaaliyhtälön ratkaisufunktioiden kuvaajat kulkevat.



Kuva 34. Suuntakenttä yhtälölle $y' = \cos^2(y)$, jonka ratkaisuja ovat esimerkiksi funktiot muotoa $y(x) = \arctan(x+C)$. Koordinaatistoon on hahmoteltu kolme ratkaisukäyrää, jotka näyttävät seuraavan nuolia.

8.4 Lineaariset differentiaaliyhtälöt

Differentiaaliyhtälöä sanotaan lineaariseksi, jos sen ratkaisujen joukolla on lineaarinen rakenne, eli jos y_1 ja y_2 ovat molemmat ratkaisuja niin myös $c_1 y_1 + c_2 y_2$ on ratkaisu kaikilla vakioilla c_1, c_2 . Esimerkkejä lineaarisista ensimmäisen ja toisen asteen differentiaaliyhtälöistä ovat

$$y' + x^2 y = 0 \quad \text{ja} \quad 2y'' + x y' - y = 0.$$

Lineaarisen yhtälön voi tunnistaa aina siitä, että ratkaisu y ja sen derivaatat $y', y'', \dots$ esiintyvät yksinään, ja niiden edessä on kertoimena joku x :n funktio. Lineaariset differentiaaliyhtälöt ovat varsin tärkeä erikoistapaus, ja niiden ratkaisemiseen on tehokkaita tekniikoita, joihin perehdymme nyt tarkemmin.

8.5 Ensimmäisen asteen lineaarinen yhtälö

Yleinen ensimmäisen asteen lineaarinen differentiaaliyhtälö on yhtälö muotoa

$$a_1(x) y' + a_0(x) y = 0, \tag{35}$$

missä a_0 ja a_1 ovat annettuja x :n funktioita. Oletetaan kuitenkin, että $a_1(x) \neq 0$, sillä jos a_1 :llä on nollakohtia niin yhtälön tarkastelu voidaan tehdä erikseen nollakohdissa ja niiden ulkopuolella. Siispä funktiolla a_1 voidaan jakaa puolittain ja saattaa yhtälö muotoon

$$y' + a(x)y = 0, \quad (36)$$

missä $a = a_0/a_1$.

Huomataan, että tämä yhtälö on tosiasiaassa myös separoituva yhtälö, sillä se voidaan palauttaa tilanteeseen (30) valitsemalla $f(x) = -a(x)$ ja $g(y) = y$. Voidaan myös huomata, että $1/g$:n integraalifunktio G on funktio $G(y) = \ln y$, joten ratkaisu y tulee esiintymään eksponenttifunktiona jostain toisesta funktiosta. Tämä tarjoaa meille myös toisen tavan ratkaista yhtälö (36), jonka käymme nyt läpi.

Idea on kertoa yhtälö puolittain tekijällä $e^{A(x)}$, missä $A'(x) = a(x)$, eli A on jokin kertoimelle a löydetty integraalifunktio. Tästä on se hyöty, että nyt voidaan soveltaa tulon derivointikaavaa käänteiseen suuntaan, ja huomata, että yhtälö

$$e^{A(x)} y' + a(x) e^{A(x)} y = 0$$

palautuu muotoon

$$\frac{d}{dx} (e^{A(x)} y) = 0.$$

Tästä on nyt helppo ratkaista, että $e^{A(x)} y = C$ jollakin vakiolla C , joten yleinen yhtälön (36) ratkaisu saadaan siis kaavasta

$$y(x) = C e^{-A(x)}.$$

Epähomogeeninen lineaarinen yhtälö

Yhtälöä muotoa

$$y' + a(x)y = b(x) \quad (37)$$

sanotaan **epähomogeeniseksi lineaariseksi yhtälöksi**. Tämä viittaa siihen, että yhtälö on lineaarinen termiä $b(x)$ lukuun ottamatta, ja oleellista on, että tämä termi on puhtaasti x :n funktio. Tällaisen yhtälön ratkaisemiseen voidaan soveltaa samaa ideaa kuin äsken käsittelemäämme homogeeniseen tapaukseen (eli tapaus $b \equiv 0$). Kerrotaan siis puolittain funktiolla $e^{A(x)}$, missä A on kertoimen a integraalifunktio, ja yhtälö palautuu muotoon

$$e^{A(x)} y' + a(x) e^{A(x)} y = b(x) e^{A(x)}$$

eli

$$\frac{d}{dx} (e^{A(x)} y) = b(x) e^{A(x)}.$$

Tästä voidaan nyt ratkaista, että yleinen ratkaisu epähomogeeniselle yhtälölle (37) löytyy kaavasta

$$y(x) = e^{-A(x)} \left(\int b(x) e^{A(x)} dx + C \right), \quad (38)$$

missä $\int b(x) e^{A(x)} dx$ merkitsee jotakin integraalifunktiota funktiolle $b(x) e^{A(x)}$, ja C on vapaavalintainen vakio.

Esimerkki. Yhtälön

$$y' + 2y = e^{-x}$$

voi ratkaista seuraavasti. Huomataan, että yhtälö on jo epähomogeenisen lineaarisen yhtälön muodossa (37), missä $a(x) = 2$, jonka yksinkertaisin integraalifunktio on $A(x) = 2x$. Kerrotaan siis puolittain termillä $e^{A(x)} = e^{2x}$, jolloin saadaan

$$\begin{aligned} e^{2x} y' + 2e^{2x} y &= e^{2x} e^{-x} \\ \frac{d}{dx} (e^{2x} y) &= e^x && \parallel \int dx \\ e^{2x} y &= e^x + C && \parallel : e^{2x} \\ y &= e^{-2x} (e^x + C). \end{aligned}$$

8.6 Toisen asteen vakiokertoiminen lineaarinen yhtälö

Toisen asteen differentiaaliyhtälössä esiintyy tuntematon funktio, sen toinen derivaatta ja mahdollisesti myös ensimmäinen derivaatta. Lineaarisen toisen asteen differentiaaliyhtälön ratkaiseminen eksplisiittisesti on vaikeampaa kuin ensimmäisen asteen tapauksessa, mutta tietyissä erikoistapauksissa se onnistuu myös. Eräs tärkeä tapaus on vakiokertoiminen lineaarinen yhtälö, eli yhtälö muotoa

$$ay'' + ay' + cy = 0, \quad a \neq 0. \quad (39)$$

Tämän yhtälön ratkaisemista varten tehdään valistunut arvaus $y(x) = e^{\lambda x}$ jollakin vakiolla λ . Sijoittamalla tämä arvaus yhtälöön (39) saadaan

$$(a\lambda^2 + b\lambda + c) e^{\lambda x} = 0,$$

mistä huomataan, että vakio λ on niin sanotun **karakteristisen yhtälön**

$$a\lambda^2 + b\lambda + c = 0 \quad (40)$$

eräs ratkaisu. Toisen asteen yhtälöllä on tyypillisesti kaksi ratkaisua, joten tämä sijoitus vaikuttaa erittäin lupaavalta. Käydään kuitenkin yksityiskohtaisesti mahdolliset tapaukset läpi.

Tapaus 1: Diskriminantti $D = b^2 - 4ac$ on positiivinen.

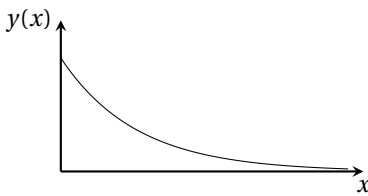
Tunnetusti diskriminantti D kertoo toisen asteen yhtälön ratkaisujen luonteen, ja tapauksessa $D > 0$ saamme ratkaisukaavasta kaksi reaalista ratkaisua

$$\lambda_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{ja} \quad \lambda_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

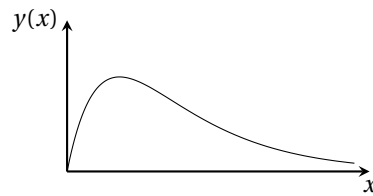
Siispä funktiot $y_1(x) = e^{\lambda_1 x}$ ja $y_2(x) = e^{\lambda_2 x}$ ovat yhtälön (39) kaksi ratkaisua, ja koska yhtälö on lineaarinen, myös funktio

$$y(x) = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x}$$

on yhtälön ratkaisu kaikilla vakioilla C_1 ja C_2 . Tämä käsittää itse asiassa kaikki yhtälön (39) mahdolliset ratkaisut. Emme perustele tätä vielä sen tarkemmin, mutta on luontevaa, että toisen asteen yhtälön ratkaisussa on tasan kaksi vapaavalintaista parametria C_1 ja C_2 .



$$y(x) = 1,5 \cdot e^{-x}$$



$$y(x) = 5 \cdot e^{-x} - 5 \cdot e^{-2x}$$

Tapaus 2: Diskriminantti on negatiivinen.

Tässä tapauksessa yhtälöllä (40) ei ole reaalisia ratkaisuja, mutta sillä on kaksi erisuuruista kompleksista ratkaisua, jotka saadaan jälleen ratkaisukaavasta muodossa

$$\lambda_1 = \frac{-b + i\sqrt{4ac - b^2}}{2a} \quad \text{ja} \quad \lambda_2 = \frac{-b - i\sqrt{4ac - b^2}}{2a}.$$

Mikäli hyväksymme kompleksiarvoiset ratkaisut yhtälölle (39), saadaan samoin kuin edellisessä tapauksessa yleinen ratkaisu kaavasta

$$y(x) = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x}. \quad (41)$$

Mikäli etsimme puhtaasti reaalisia ratkaisuja, ne voidaan löytää seuraavasti. Muistetaan trigonometristen funktioiden yhteys kompleksiseen eksponenttifunktioon, eli kaavat

$$\sin(x) = \frac{1}{2i} (e^{ix} - e^{-ix}) \quad \text{ja} \quad \cos(x) = \frac{1}{2} (e^{ix} + e^{-ix}).$$

Näitä kaavoja hyödyntämällä voidaan päästä eteenpäin seuraavasti. Kirjoitetaan ensin vakiot λ_1 ja λ_2 muodossa

$$\lambda_1 = \alpha + i\beta \quad \text{ja} \quad \lambda_2 = \alpha - i\beta, \quad \text{missä} \quad \alpha = \frac{-b}{2a}, \beta = \frac{\sqrt{4ac - b^2}}{2a}.$$

Sijoitetaan nyt yhtälöön (41) ensin vakiot $C_1 = \frac{1}{2i}$ ja $C_2 = -\frac{1}{2i}$, joilla saadaan eräs ratkaisu

$$\begin{aligned} u_1(x) &= \frac{1}{2}e^{\lambda_1 x} - \frac{1}{2}e^{\lambda_2 x} = \frac{1}{2}e^{\alpha x + i\beta x} - \frac{1}{2}e^{\alpha x - i\beta x} \\ &= e^{\alpha x} \frac{1}{2i} (e^{i\beta x} - e^{-i\beta x}) = e^{\alpha x} \sin(\beta x). \end{aligned}$$

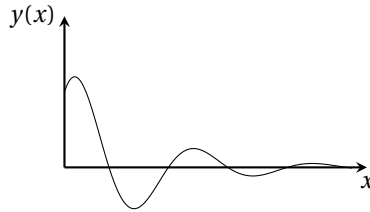
Vastaavalla laskulla sijoituksella $C_1 = \frac{1}{2}$ ja $C_2 = \frac{1}{2}$ saadaan toinen ratkaisu

$$u_2(x) = e^{\alpha x} \cos(\beta x).$$

Nämä kaksi ratkaisua u_1 ja u_2 ovat selvästi reaaliarvoisia. Tämä tarjoaa meille uuden tavan esittää yhtälön (39) ratkaisut muodossa

$$y(x) = C_1 e^{\alpha x} \sin(\beta x) + C_2 e^{\alpha x} \cos(\beta x), \quad (42)$$

missä C_1 ja C_2 ovat jälleen vakioita, jotka voidaan valita joko reaalisiksi tai kompleksisiksi riippuen siitä, millaisia ratkaisuja etsitään.



$$y(x) = (\cos 4x + \sin 4x) \cdot e^{-x}$$

Tapaus 3: Diskriminantti on nolla.

Tässä tapauksessa yhtälöllä (40) on kaksinkertainen nollakohta pisteessä

$$\lambda = \frac{-b}{2a}. \quad (43)$$

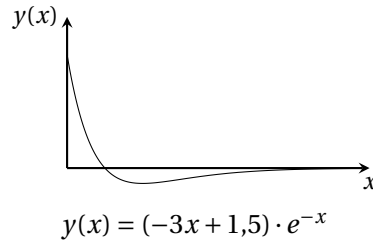
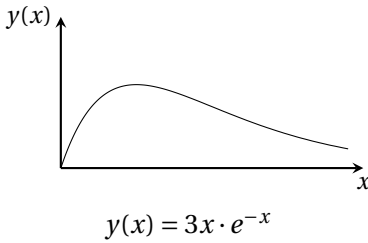
Tässä tapauksessa toki funktio $y_1(x) = e^{\lambda x}$ on eräs ratkaisu alkuperäiselle yhtälölle (39), mutta toivoisimme että löytyisi myös toinen ratkaisu. Käy itseasiassa niin, että funktio $y_2(x) = x e^{\lambda x}$ ratkaisee yhtälön myös, sillä laskemalla saadaan

$$\begin{aligned} a y_2'' + b y_2' + c y_2 &= a (\lambda^2 x e^{\lambda x} + 2\lambda e^{\lambda x}) + b (\lambda x e^{\lambda x} + e^{\lambda x}) + c x e^{\lambda x} \\ &= (a\lambda^2 + b\lambda + c) x e^{\lambda x} + (2a\lambda + b) e^{\lambda x}. \end{aligned}$$

Viimeisessä lausekkeessa molemmat sulkeiden sisällä olevat lausekkeet ovat nollia. Ensimmäinen siksi, että λ oli karakteristisen yhtälön ratkaisu, ja toinen kaavan (43) nojalla. Siispä y_2 todella on differentiaaliyhtälömmme ratkaisu, ja yleiselle ratkaisulle voidaan nyt muodostaa kaava

$$y(x) = C_1 e^{\lambda x} + C_2 x e^{\lambda x}, \quad (44)$$

missä C_1 ja C_2 ovat jälleen vakioita.

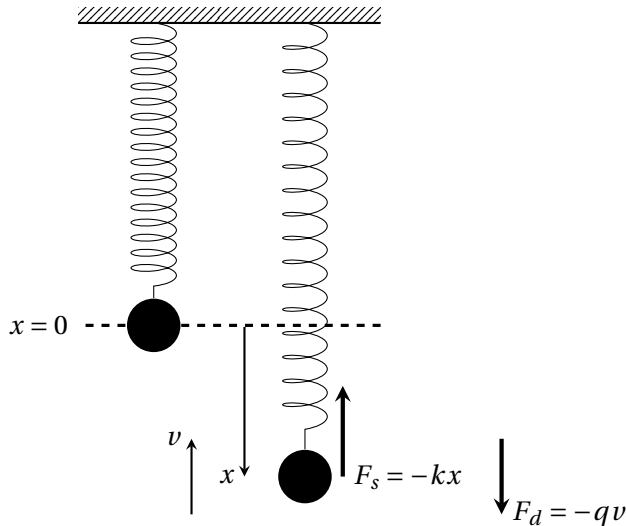


Esimerkki. Ratkaistaan differentiaaliyhtälö

$$y'' = -y.$$

Karakteristinen yhtälö on tässä tapauksessa muotoa $\lambda^2 + 1 = 0$, jonka ratkaisut saadaan kaavasta $\lambda = \pm i$. Ratkaisutapauksen 2 merkinnöillä saadaan $\alpha = 0$ ja $\beta = 1$, joten sijoittamalla yhtälöön (42) huomataan, että yleinen ratkaisu on muotoa $y(x) = C_1 \sin(x) + C_2 \cos(x)$.

Esimerkki 2. Vaimennettu harmoninen värähtelijä. Tarkistellaan katosta jousen varassa roikkuvaa kappaletta. Sovitaan, että kappale liikkuu ainoastaan ylös ja alas, ja kuvataan sen sijaintia ajasta riippuvalla funktiolla $x(t)$. Sovitaan myös, että tasapainoasema löytyy kohdasta $x = 0$, eli $x(t)$ kuvaa poikkeamaa tasapainoasemasta.



Joka hetkellä kun kappale ei ole tasapainossa, siihen kohdistuu kaksi voimaa. Ensimmäinen on jousen aiheuttama voima F_s , joka on **Hooken lain** mukaan verrannollinen kappaleen poikkeamaan tasapainoasemasta, eli $F_s = -kx(t)$, missä $k > 0$ on jousivakio. Voiman suunta on vastakkainen siirtymään x nähden, sillä jousi pyrkii palauttamaan kappaletta takaisin tasapainoon.

Toinen voima on väliaineen vastus F_d . Tyypillisesti se on verrannollinen kappaleen nopeuden neliöön, mutta jos väliaine on riittävän paksua ja nopeus riittävän pieni,

vastus on suoraan verrannollinen nopeuteen. Oletetaan vaikkapa, että kuula on upotettu suureen purkkiin juoksevaa hunajaa. Tällöin voidaan olettaa $F_d = -qv$, missä $q > 0$ on vakio ja $v = x'(t)$ on kappaleen nopeus. Väliaineen vastuksen suunta on päinvastainen kuin nopeus.

Newtonin toisen lain mukaan $\sum F = ma = mx''(t)$, joten ajanhetkellä t saadaan yhtälö $mx''(t) = F_s + F_d = -kx(t) - qx'(t)$. Kappaleen liikettä esittää siis toisen asteen vakiokertominen differentiaaliyhtälö

$$0 = mx''(t) + qx'(t) + kx(t).$$

Tätä yhtälöä on luonnollista tutkia alkuarvo-ongelmana, johon asetetaan alkuarvoiksi ajanhetkellä $t = 0$ kappaleen alkusijainti $x(0) = x_0$ ja alkunopeus $x'(0) = v_0$. Riippuen vakioiden m, q , ja k arvoista, voidaan tätä yhtälöä ratkaistaessa päätyä mihin tahansa kolmesta aiemmin esitetystä ratkaisutilanteesta.

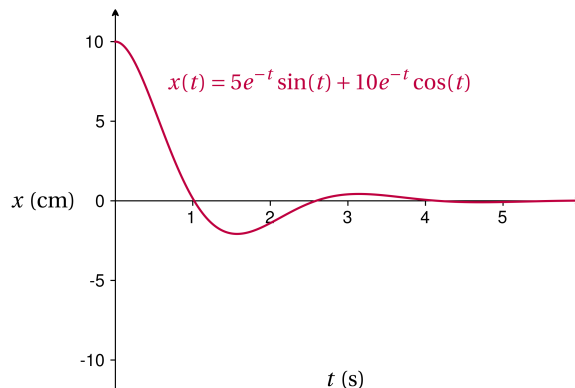
Tutkitaan yksittäistä esimerkkiä, jossa vaikkapa $m = 1$ kg, $q = 2$ kg·s, ja $k = 5$ kg·s². Tällöin karakteristinen yhtälö on muotoa $\lambda^2 + 2\lambda + 5 = 0$, jonka diskriminantti $D = 2^2 - 4 \cdot 1 \cdot 5 = -16$ on negatiivinen. Saadaan kaksi kompleksista ratkaisua

$$\lambda = \frac{-2 \pm \sqrt{-16}}{2} = -1 \pm 2i.$$

Ratkaisutapauksen 2 merkinnöillä saadaan $\alpha = -1$ ja $\beta = 2$, joten yleinen ratkaisu on muotoa

$$x(t) = C_1 e^{-t} \sin(2t) + C_2 e^{-t} \cos(2t).$$

Alkuarvoista $x_0 = 10$ cm ja $v_0 = 0$ cm/s saadaan yhtälöt $C_2 = 10$ ja $2C_1 - C_2 = 0$, eli $C_1 = 5$. Hahmotellaan vielä ratkaisun kuvaaja.



Kuva 35. Hunajaan upotetun harmonisen värähtelijän poikkeama tasapainoasemasta.

Harjoitustehtäviä

Tehtävä 63. Ratkaise alkuarvo-ongelma $y' = xe^{-y}$, $y(0) = 1$.

Vihje s. 162 →

Toinen vihje s. 178 →

Tehtävä 64. Hahmottele differentiaaliyhtälön

$$y' = y^2 x$$

suuntakenttä.

Vihje s. 162 →

Toinen vihje s. 178 →

Tehtävä 65. Selvitä, onko yhtälö $y' = xy + x + y + 1$ separoituva, ja etsi kaikki sen ratkaisut.

Vihje s. 162 →

Toinen vihje s. 178 →

Tehtävä 66. Määritä funktio y , joka toteuttaa pisteillä $x > 0$ sekä differentiaaliyhtälön $y' = y \cdot (\frac{1}{x} - 2)$ että differentiaaliyhtälön $y' + 2y = e^{-2x}$.

Vihje s. 162 →

Tehtävä 67. Kappaleen jäähtymistä voidaan mallintaa differentiaaliyhtälöllä

$$T'(t) = -k(T(t) - M),$$

missä T kuvaa kappaleen lämpötilaa ajanhetkellä t , ympäröivän ilman lämpötila on M , ja vakiokerroin $k > 0$ riippuu kappaleesta.

Eräässä kokeessa havaittiin, että 200-asteinen metallilevy jäähdyi 100 asteeseen kymmenessä minuutissa, kun se oli 10-asteisessa ulkoilmassa. Määritä tälle kappaleelle yhtälön kerroin k , kun ajan yksikkönä on sekunti.

Vihje s. 162 →

Toinen vihje s. 179 →

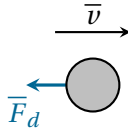
Tehtävä 68. Edellinen koe toistettiin uudestaan seuraavana päivänä kun sää oli viileämpi, ja tällä kertaa sama kappale jäähdyi 200 asteesta 100 asteeseen yhdeksässä minuutissa. Selvitä, mikä oli uusi ulkoilman lämpötila.

Vihje s. 162 →

Toinen vihje s. 179 →

Tehtävä 69. Ilmanvastus. Lukiofysiikassa ilmanvastus arvioidaan usein nolllaksi. Tutkitaan nyt mahdollisimman yksinkertaista tilannetta, jossa ilmanvastus on tärkeä, ja oleellisesti ottaen ainoa voima: Sysätään avaruusasemalla tennispallo liikkeelle alkuvauhdilla v_0 , jolloin palloon vaikuttaa vain ilmanvastus.

Todella pieniä nopeuksia lukuun ottamatta ilmanvastuksen F_d suuruus on likimain verrannollinen vauhdin toiseen potenssiin v^2 . Kullakin hetkellä ilmanvastus on vastakkaissuuntainen nopeusvektorin $\vec{v}$ kanssa, eli jos liikkeen suunta valitaan positiiviseksi suunnaksi, saadaan $F_d = -qv^2$, missä q on positiivinen vakio.



- Ratkaise sopivalla differentiaaliyhtälöllä pallon vauhti $v(t)$ ajan funktiona, kun $v(0) = v_0$.
- Käytä arvoja $m = 0,057 \text{ kg}$, $q = 0,0012 \frac{\text{kg}}{\text{m}}$, ja $v_0 = 1,0 \frac{\text{m}}{\text{s}}$. Mikä on pallon vauhti, kun on kulunut aika $t = 5 \text{ s}$?
- Kuinka pitkän matkan pallo kulkee aikavälillä $0 \text{ s} \dots 5 \text{ s}$?

Ensimmäisessä vihjeessä on opastusta fysiikkaan, toisessa differentiaaliyhtälön ratkaisuun.

Vihje s. 163 →

Toinen vihje s. 179 →

Tehtävä 70. Osoita, että alkuarvo-ongelmalla

$$y' = 3y^{\frac{2}{3}}, \quad y(0) = 0$$

on a) enemmän kuin yksi b) äärettömän monta eri ratkaisua.

Vihje s. 163 →

Toinen vihje s. 179 →

Tehtävä 71.

Tutkitaan alkuarvo-ongelman $y' = e^{x^2} (1 - e^{y^2 \cos(y)})$, $y(0) = 1$ ratkaisua. Osoita, että ratkaisulle pätee $-\pi/2 < y(x) < \pi/2$ kaikilla muuttujan x arvoilla.

Vihje s. 163 →

Toinen vihje s. 179 →

Tehtävä 72. Etsi differentiaaliyhtälön

$$y'' - 4y' + 4y = 0$$

yleinen ratkaisu, ja selvitä myös ratkaisu alkuarvo-ongelmaan $y(0) = 2$, $y'(0) = 5$.

Vihje s. 163 →

Toinen vihje s. 180 →

Tehtävä 73. Eräs tyypillinen sovelluksissa esiintyvä toisen asteen lineaarinen differentiaaliyhtälö on

$$x^2 y'' + ax y' + by = 0,$$

missä a ja b ovat vakioita. Fysiikassa usein esiintyvä **Laplacen yhtälö** muutetaan usein napakoordinaatteihin, jolloin päädytään ratkaisemaan tällaista yhtälöä.

Yhtälö ei kuitenkaan ole vakiokertoiminen, joten sitä ei voida ratkaista aivan samalla tavalla kuin aiemmin. Tee sen sijaan sijoitus $y(x) = x^r$, missä r on määritettävä vakio, ja selvitä voiko tätä soveltaa yhtälön ratkaisemiseen.

Ratkaise yhtälön yleinen ratkaisu myös tarkasti tapauksessa $a = 3$ ja $b = -3$. Voit olettaa, että yhtälöllä on vain kaksi lineaarisesti riippumatonta ratkaisua.

Vihje s. 163 →

Toinen vihje s. 180 →

9 Differentiaaliyhtälön numeerinen ratkaiseminen

Olemme tähän mennessä esittäneet useita tapoja ratkaista differentiaaliyhtälöitä eksplisiittisesti, eli tapoja löytää ratkaisufunktiolle tutuista alkeisfunktioista koostuva kaava. Kuitenkin suurimmassa osassa käytännön sovelluksista yhtälöt ovat sellaista muotoa, että eksplisiittistä ratkaisua ei yksinkertaisesti ole mahdollista löytää. Joskus taas ratkaisu saadaan suljettuun muotoon, mutta sen arvoja on liian työlästä laskea jopa tietokoneella.

Näistä syistä differentiaaliyhtälöiden ratkaisemiseen sovelletaan käytännön tilanteissa lähes aina numeerisia menetelmiä. Numeeristen menetelmien tavoitteena on antaa likiarvoja ratkaisufunktion todellisille arvoille, ja näiden avulla voidaan hahmottaa ratkaisufunktion kuvaajaa sekä sen ominaisuuksia.

9.1 Eulerin menetelmä



Leonhard Euler (1707–1783)

Ensimmäinen esittelemämme approksimaatio on nimeltään **Eulerin menetelmä**, jonka etuna on ennen kaikkea yksinkertaisuus. Perusmuodossaan Eulerin menetelmä soveltuu ensimmäisen asteen differentiaaliyhtälöihin, eli tilanteeseen, jossa tutkitaan yhtälöitä muotoa

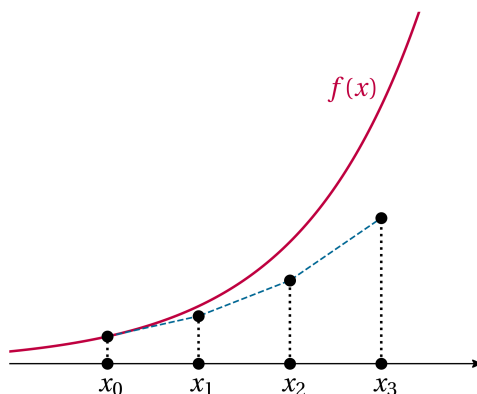
$$y' = h(x, y), \quad (45)$$

missä h on annettu funktio. Pyritään etsimään approksimaatio ratkaisulle annetun alkuarvon $y(x_0) = y_0$ läheisyydessä.

Eulerin menetelmän ensimmäisessä askeleessa approksimoimme todellista ratkaisua $y(x)$ sen kuvaajan pisteeseen (x_0, y_0) piirretyllä tangenttisuoralla. Tangentin kulmakerroin lasketaan yhtälön (45) avulla: derivaatta alkuarvopisteessä (x_0, y_0) saadaan sijoittamalla $y'(x_0) = h(x_0, y_0)$. Tangenttisuoran yhtälö on

$$T_{x_0}(x) = h(x_0, y_0)(x - x_0) + y_0.$$

Kun x on lähellä alkuarvopistettä x_0 , tämä tangenttisuora $T_{x_0}(x)$ tarjoaa jo alustavan approksimaation todellisen ratkaisufunktion $y(x)$ arvoista. Seuraavaksi pyrimme jatkamaan tätä approksimaatiota kauemmas pisteestä x_0 ottamalla ensin pienen askeleen tangentin suunnassa, laskemalla derivaatan uudessa pisteessä, ja toistamalla tätä prosessia.



Kuva 36. Eulerin menetelmässä yhtälön ratkaisua approksimoidaan muodostamalla jono lineaarisia approksimaatioita.

Valitaan siis jokin pieni positiivinen luku Δx , jota kutsumme **askelpituudeksi**. Kun otetaan x -akselia pitkin askelpituuden mittainen harppaus pisteestä x_0 , päädytään uuteen pisteeseen $x_1 = x_0 + \Delta x$. Aiemmin löytämämme tangenttisuora pisteessä x_0 antaa meille tässä pisteessä ratkaisulle approksimaation $y(x_1) \approx h(x_0, y_0)\Delta x + y_0$; merkitään tätä approksimaatioarvoa $y_1 = h(x_0, y_0)\Delta x + y_0$. Nyt voidaan toistaa tähän mennessä tehty prosessi uudestaan lähtien pisteestä (x_1, y_1) . Tässä pisteessä saadaan uusi approksimaatiosuora

$$T_{x_1}(x) = h(x_1, y_1)(x - x_1) + y_1,$$

ja vastaavasti tätä suoraa pitkin voidaan ottaa x -suunnassa Δx :n pituinen askel ja päädytään seuraavaan pisteeseen (x_2, y_2) . Prosessia voidaan nyt jatkaa rekursiivi-

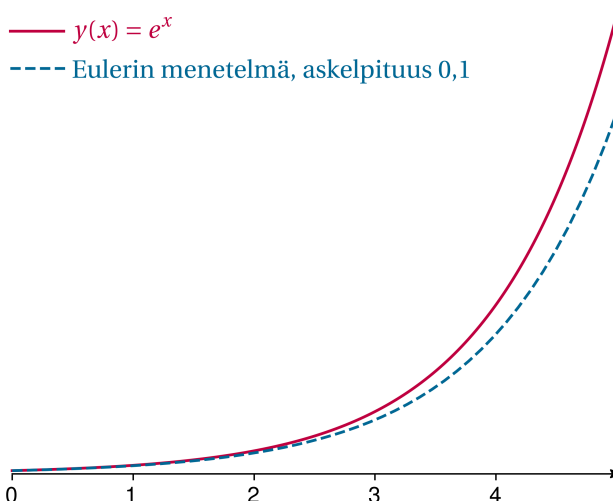
sesti, ja saadaan jono pisteitä (x_n, y_n) , $n = 1, 2, \dots$ joille löytyy rekursiokaavat

Eulerin menetelmä

$$\begin{cases} x_n = x_0 + n\Delta x \\ y_n = y_{n-1} + h(x_{n-1}, y_{n-1})\Delta x \end{cases} \quad (46)$$

kun $n = 1, 2, \dots$

Näiden pisteiden kautta kulkeva murtoviiva tarjoaa nyt arvion siitä, miltä ratkaisufunktion kuvaaja näyttää alkuarvopisteen läheisyydessä. Arvio tyypillisesti huononee, kun mennään kauemmaksi alkuarvosta x_0 .



Kuva 37. Eulerin menetelmä yhtälölle $y' = y$ alkuarvolla $y(0) = 1$, jonka eksakti ratkaisu on $y(x) = e^x$. Askelpituus on $\Delta x = 0,1$.

Numeerisissa menetelmissä on tyypillistä, että mitä enemmän laskenta-aikaa käytetään, sen tarkempi approksimaatio voidaan saavuttaa. Eulerin menetelmässä tämä tapahtuu pienentämällä askelväliä Δx , jolloin tietyllä välillä ratkaisun ennustaminen vaatii enemmän laskemista, mutta tuottaa paremman approksimaation.

Esimerkki, Eulerin menetelmä. Palataan edellisessä luvussa esittämäämme jänispopulaatiomalliin, ja tehdään siitä vähän hienovaraisempi. Aiemmassa mallissa jänispopulaation $y(t)$ kasvulla ei ollut rajoja, mutta sovitaan nyt, että ympäristö ei pysty ylläpitämään liian suuria populaatiota. Kuvitellaan, että ympäristöstä johtuvista tekijöistä teoreettinen maksimipopulaatio saadaan kantokykyfunktioista $K(t)$, joka on mallikohtainen ja voi tässä riippua myös ajasta t .

Lisätään populaatiomallin yhtälöön (29) kasvukerroin $1 - y(t)/K(t)$, joka hidastaa populaation kasvua kun populaation koko $y(t)$ lähestyy maksimipopulaatiota $K(t)$

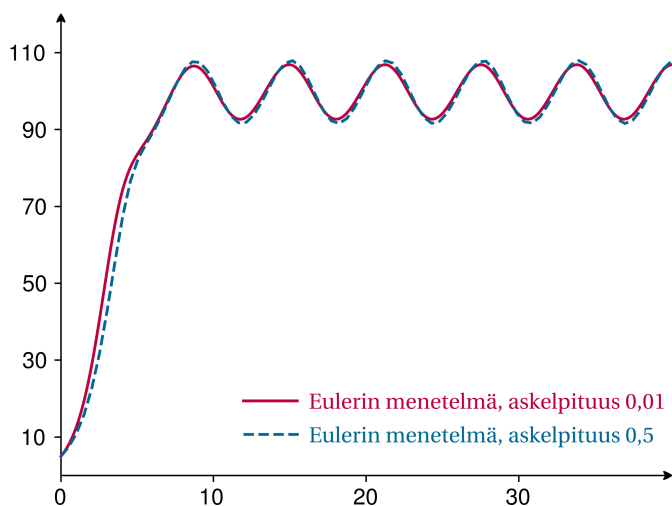
ja kääntää kasvun negatiiviseksi, jos populaation koko ylittää arvon $K(t)$. Saadaan uusi differentiaaliyhtälö

$$y'(t) = ky(t) \left(1 - \frac{y(t)}{K(t)} \right).$$

Tämä yhtälö tunnetaan **aikariippuvaisena logistisena yhtälönä**. Jos kantokykyfunktio K on vakio, niin yhtälö on ratkaistavissa eksplisiittisesti separoituvana yhtälönä, mutta jos K on vähänkin monimutkaisempi muuttujan t funktio, on helppo kuvitella, että tällainen eksplisiittinen ratkaiseminen ei onnistu.

Havainnollistetaan nyt mallia kannatusfunktioilla $K(t) = 100 + 10 \sin(t)$. Tässä sini-funktio tuo ympäristön suotuisuuteen heilahtelua ajan suhteen, mikä voisi esimerkiksi alkeellisella tavalla kuvastaa vuodenaikoja. Selvitetään Eulerin menetelmällä mitä käy, kun $k = 1$ ja alkupopulaatio on $y(0) = 5$.

Tutkitaan kahta askelpituutta $\Delta x = 0,01$ ja $\Delta x = 0,5$. Oheisessa kuvassa näkyy näiden askelpituuksien tuottamat ratkaisuaapproksimaatiot.



Kuva 38. Eulerin menetelmä kahdella eri askelpituudella.

Molemmissa approksimaatioissa näkyy jo, että populaation koko pysyy rajoitettuna kuten haluttiin, ja se heilahtelee myös kannatusfunktion heilahtelun mukana. Pienempi askelpituus luonnollisesti tarkentaa arviota entisestään.

9.2 Runge–Kutta-menetelmä

Differentiaaliyhtälöiden ratkaisujen approksimoimiseen sovelletaan käytännössä yleensä hienostuneempaa versiota Eulerin menetelmästä nimeltä **Runge–Kutta-menetelmä**, tai Rungen–Kuttan menetelmä, sillä menetelmän keksivät saksalaiset

Carl Runge (1856–1927) ja Martin Kutta (1867–1944). Kuten Eulerin menetelmässä, tavoite on approksimoida ratkaisufunktiota askel askeleelta. Ero näissä kahdessa menetelmässä on se, että siinä missä Eulerin menetelmässä seuraavan approksiimaatiopisteen y -koordinaatti y_{n+1} laskettiin suoraan kaavasta

$$y_{n+1} = h(x_n, y_n) \Delta x + y_n,$$

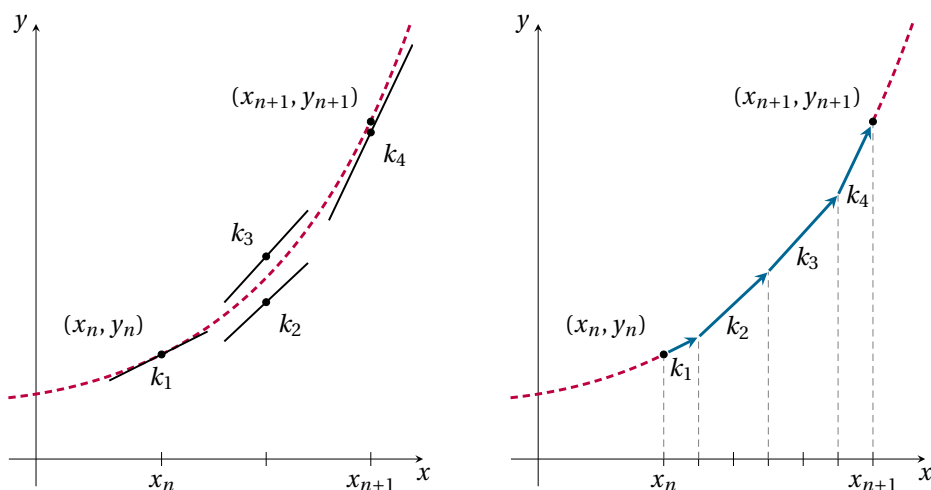
niin Runge–Kutta-menetelmässä lasketaan ensin välituloksena derivaatalle arvo rekursiivisesti neljässä eri paikassa:

$$\begin{aligned} k_1 &= h(x_n, y_n) \\ k_2 &= h\left(x_n + \frac{\Delta x}{2}, y_n + k_1 \cdot \frac{\Delta x}{2}\right) \\ k_3 &= h\left(x_n + \frac{\Delta x}{2}, y_n + k_2 \cdot \frac{\Delta x}{2}\right) \\ k_4 &= h(x_n + \Delta x, y_n + k_3 \cdot \Delta x). \end{aligned} \tag{47}$$

Kokonaisessa askeleessa lopulta käytettävä kulmakerroin lasketaan näiden neljän kulmakertoimen painotettuna keskiarvona, ja seuraavaksi y -koordinaatiksi y_{n+1} saadaan

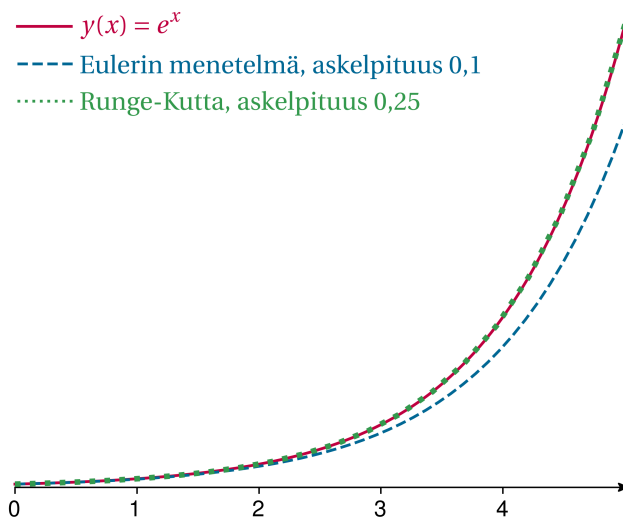
$$y_{n+1} = y_n + \frac{k_1 + 2k_2 + 2k_3 + k_4}{6} \cdot \Delta x.$$

Tiivistettynä Runge–Kutta-menetelmässä idea on tuottaa tarkempi approksimaatio ratkaisun arvojen muutokselle laskemalla funktion h arvoja useammassa pisteessä. Näitä arvoja sopivasti painottamalla saadaan lopulta hienovaraisempi arvio todellisen ratkaisufunktion käyttäytymisestä.



Kuva 39. Runge–Kutta-menetelmän yksi askel. Tarkka ratkaisu katkoviivalla.

Laskentatehollisesti Runge–Kutta on noin neljä kertaa niin kallis kuin Eulerin menetelmä samalla askelpituudella, mutta se tuottaa paljon tarkempia tuloksia.



Kuva 40. Yhtälön $y' = y$ ratkaisun approksimaatio Eulerin ja Runge–Kutta-menetelmillä. Eulerin menetelmä on kauempana oikeasta ratkaisusta, vaikka käytetty askelpituus oli pienempi.

Numeeristen menetelmien virhearviot

Puhutaan seuraavaksi edellä esiteltyjen menetelmien tarkkuudesta. Oletamme tässä, että yhtälön määrittelevä funktio h käyttäytyy suhteellisen hyvin, mikä tässä yhteydessä tarkoittaa, että sillä on olemassa molempien muuttujien suhteen korkeampienkin asteiden derivaatat. Molemmissa esittelemissämme menetelmissä tapahtuu jokaisella rekursioaskeleella pieni virhe, joka tyypillisesti kasaantuu aina vain suuremmaksi mitä pidemmälle mennään. Jos kuitenkin yksittäisen askeleen virhe on riittävän pieni, niin kokonaisvirhettä voidaan myös hallita.

Eulerin menetelmä perustuu joka askeleella puhtaasti kulmakertoimen eli ensimmäisen asteen derivaatan arvioimiseen. Siinä on siis itse asiassa kyse todellisen ratkaisun vertaamisesta ensimmäisen asteen Taylorin polynomiin joka iteraatioaskeleella. Tästä syystä yksittäisessä askeleessa tapahtuva virhe Eulerin menetelmässä on tyypillisesti suuruusluokkaa $(\Delta x)^2$, ja kokonaisvirhe kiinteän mittaisella välillä on suunnilleen Δx . Esitämme tarkan perustelun tälle liitteessä F (s. 150).

Runge–Kutta-menetelmän tarjoama arvio on kuitenkin huomattavasti parempi. Virhearviointi on tässä tapauksessa hieman hienovaraisempaa, emmekä mene tekniisiin yksityiskohtiin. Syy menetelmän tehokkuudelle perustuu oleellisesti ottaen siihen, että ottamalla sopivasti painotettu keskiarvo menetelmässä lasketuista

lukuista k_i saadaan ratkaisufunktion Taylorin kehitelmästä kumottua ensimmäiset neljä derivaattatermiä. Tämä johtaa siihen, että Runge–Kutta-menetelmän yhdellä askeleella tapahtuva virhe on verrannollinen askelpituuden viidenteen potenssiin $(\Delta x)^5$. Kiinteällä x -suuntaisella välillä tapahtuva virhe tulee olemaan suuruusluokkaa $(\Delta x)^4$.

Vaikka Runge–Kutta-menetelmässä joudutaan siis laskemaan funktion h arvo neljässä eri pisteessä askelta kohden ja Eulerin menetelmässä vain yhdessä, niin virhetermi pienenee jopa neljanteen potenssiin aiemmasta.

9.3 Differentiaaliyhtälöryhmät

Reaalimaailman ilmiöissä esiintyy tyypillisesti useita eri suureita joiden kasvunopeudet tai korkeammat derivaatat riippuvat toisistaan. Tulemme havainnollistamaan tätä populaatiomallilla, jossa esiintyy jänisten lisäksi myös niitä saalistavia susia, ja missä näiden kahden populaation koot vaikuttavat toisiinsa. Tällaisia ilmiöitä varten tarvitsemme yleensä useamman differentiaaliyhtälön, joissa esiintyy enemmän kuin yksi tuntematon funktio. Eräs esimerkki ensimmäisen asteen kahden funktion differentiaaliyhtälöryhmästä voisi olla vaikkapa

$$\begin{cases} y_1' = y_2 \\ y_2' = -y_1 \end{cases},$$

missä y_1 ja y_2 ovat tuntemattomia funktioita. Eräs ratkaisu tälle yhtälölle saadaan trigonometrisista funktioista $y_1(x) = \sin x$, $y_2(x) = \cos x$.

Differentiaaliyhtälöryhmien eksplisiittiseen ratkaisemiseen on hyvin samankaltaisia menetelmiä kuin yhden tuntemattoman funktion tapauksessa, mutta ne ovat edelleen hyvin rajoitettuja ja tepsivät vain tiettyntyyppisiin yhtälöihin. Keskitymme sen sijaan mieluummin numeerisiin menetelmiin, sillä ne ovat usein hyödyllisempiä käytännössä ja useamman tuntemattoman funktion tapaus ei tuota niiden kannalta suurempia lisävaikeuksia. Selvitämme aihetta esimerkin kautta.

Populaatioesimerkki 2. Edellisessä jänisesimerkissä populaation kokoa rajoitti ympäristön kantokyky. Poistetaan nyt tämä rajoitus ja tuodaan rajoittavaksi tekijäksi petoeläimet. Mallimme tulee olemaan edelleen liian yksinkertainen kuvaamaan minkään todellisten lajien vuorovaikutussuhteita, mutta se antaa karkean yleiskuvan saalistajien ja saaliiden mahdollisesta populaatiodynamiikasta.

Kuvitellaan siis, että metsässä asuu sekä jäniksiä, joiden populaation määrää kuvaa funktio $R(t)$, että susia, joiden populaatiota kuvaa toinen funktio $W(t)$. Jänispopulaation kasvuun ajanhetkellä t vaikuttaa edelleen ainakin niiden luonnollinen syntyvyys ja kuolevuus, joiden summaa kuvaamme lausekkeella $k_1 R(t)$, missä k_1 on kasvunopeutta kuvaava vakio.

Jänispopulaation muutokseen vaikuttaa myös se, kuinka nopeasti sudet syövät

jäniksiä. Mallinamme tätä suureta lausekkeella $e_1 R(t)W(t)$, missä e_1 on sopiva positiivinen kerroin. Malli on sikäli luonteva, että tulo $R(t)W(t)$ kuvaa kaikkien mahdollisten jänis-susi-parien lukumäärää. Jokainen tällainen pari johtaa tietyllä todennäköisyydellä tapaamiseen ajanhetkellä t , ja näistä tietty osuus päättyy suden eduksi, mistä saadaan lopullinen termi $e_1 R(t)W(t)$.

Kokonaisuudessaan saamme jänisten populaation muutosta hallitsevaksi differentiaaliyhtälöksi yhtälön

$$R'(t) = k_1 R(t) - e_1 R(t)W(t).$$

Tutkitaan seuraavaksi susien populaatiota. Sudet tarvitsevat jäniksiä elääkseen, joten niiden luonnollinen syntyvyys ja kuolevuuden erotus ilman jäniksiä on negatiivinen. Mallinamme tätä terminä $-k_2 W(t)$, missä kerroin k_2 on positiivinen.

Jotta susipopulaatio pääsisi kasvuun, sudet tarvitsevat ravinnokseen jäniksiä. Tämä voidaan ottaa huomioon lisäämällä kasvunopeuteen saalistusta kuvaava termi $e_2 R(t)W(t)$, missä e_2 on sopiva positiivinen kerroin. Susien populaatiota hallitsee siis differentiaaliyhtälö

$$W'(t) = e_2 R(t)W(t) - k_2 W(t).$$

Koottuna jänisten ja susien populaatiomallia esittää differentiaaliyhtälöryhmä

$$\begin{cases} R'(t) = k_1 R(t) - e_1 R(t)W(t) \\ W'(t) = e_2 R(t)W(t) - k_2 W(t) \end{cases} \quad (48)$$

missä positiiviset vakiot k_1, k_2, e_1 ja e_2 ovat mallikohtaisia. (Tyypillisesti $e_2 < e_1$, sillä yhden saaliseläimen energia ei riitä kasvattamaan uutta saalistajaa. Toisaalta funktioiden W ja R ei tarvitse olla samassa yksikössä.)

$$\begin{array}{ccccc} \begin{array}{c} \text{[Rabbit icon]} \\ R'(t) \end{array} & = & \begin{array}{c} \text{[Rabbit icon]} \\ k_1 R(t) \end{array} & - & \begin{array}{c} \text{[Rabbit icon]} \times \text{[Wolf icon]} \\ e_1 R(t)W(t) \end{array} \\ \\ \begin{array}{c} \text{[Wolf icon]} \\ W'(t) \end{array} & = & \begin{array}{c} \text{[Rabbit icon]} \times \text{[Wolf icon]} \\ e_2 R(t)W(t) \end{array} & - & \begin{array}{c} \text{[Wolf icon]} \\ k_2 W(t) \end{array} \end{array}$$

Tämä yhtälöryhmä tunnetaan **Lotkan–Volterran peto-saalismallina** (Alfred J. Lotka, 1880–1949 ja Vito Volterra, 1860–1940). Tutkitaan nyt, miten yhtälöryhmän

ratkaisuja voidaan arvioida numeerisesti Eulerin menetelmällä. Sovitaan alkuarvoiksi $R(0) = 40$ ja $W(0) = 10$, sekä asetetaan vakioparametreiksi $k_1 = 0,1$, $k_2 = 0,1$, $e_1 = 0,02$, $e_2 = 0,01$.

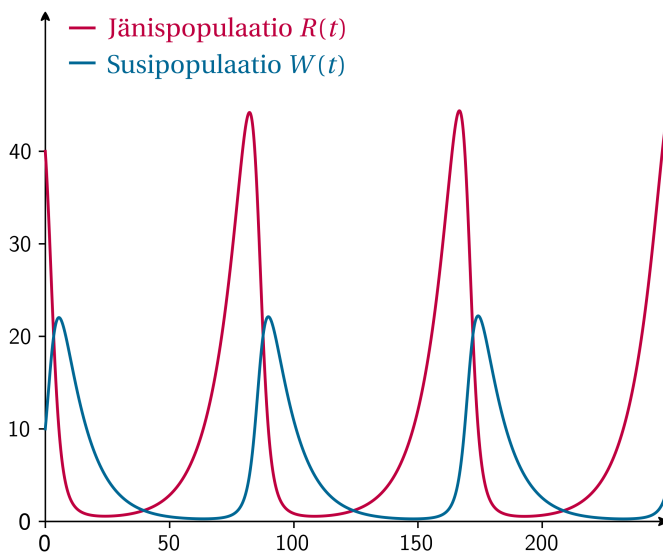
Aiemmin esittämämme idea Eulerin menetelmän soveltamiseen toimii täysin samalla tavalla kuin yhden funktion tapauksessa. Alkuhetkellä $t = 0$ voimme sijoittaa alkuarvot yhtälöön, ja saamme selville molempien populaatioiden alkuhetken kasvunopeudet $R'(0)$ ja $W'(0)$. Valitsemalla askelpituus Δt voidaan nyt ottaa yksi askel aikajanalla ja arvioida, että seuraavalla ajanhetkellä $t_1 = \Delta t$ pätee

$$R(t_1) \approx R(0) + R'(0)\Delta t \quad \text{ja} \quad W(t_1) \approx W(0) + W'(0)\Delta t.$$

Kuten yhden yhtälön tapauksessa, tästä saadaan rekursiokaavat joissa seuraavat approksimaatiokuvaajien y -koordinaatit R_{n+1} ja W_{n+1} määritetään edellisistä koordinaateista R_n ja W_n . Yhtälön perusteella rekursiokaavoiksi tulee

$$\begin{cases} R_{n+1} = R_n + (k_1 R_n - e_1 R_n W_n)\Delta t \\ W_{n+1} = W_n + (e_2 R_n W_n - k_2 W_n)\Delta t \end{cases}$$

Oheisessa kuvaajassa on esitetty tämän kaavan antamat ratkaisuapproksimaatiot (t_n, R_n) ja (t_n, W_n) askelpituudella $\Delta t = 0,01$.



Kuva 41. Jänisten ja susien populaatiomalli Eulerin menetelmällä.

Kuvaajissa ilmenee luonnollisella tavalla populaatioiden heilahtelu. Kun jäniksiä on paljon, susien kanta kasvaa voimakkaasti, ja jänisten määrä alkaa vähetä. Saaliiden määrän vähentyessä susipopulaation kasvu hidastuu ja kääntyy laskuun. Kun susien määrä on vähennyt pieneksi, jänispopulaatio alkaa taas kasvaa

nopeasti ja sykli toistuu. Tämä malli on edelleen melko alkeellinen, mutta sen enustamaa kantojen voimakasta jaksollista vaihtelua esiintyy luonnossa oikeastikin tilanteissa, joissa saalis- ja saalistajalajeja on vähän.

Käyttämämme ratkaisumenetelmä soveltuu myös monimutkaisempiin malleihin ja mallinnettaviin ilmiöihin, mutta kasautuvan virheen vuoksi Eulerin menetelmä joutuu usein ongelmiin pitkien simulaatioiden kanssa. Esimerkiksi edellisessä jänisten ja susien mallissa populaatioiden maksimikoot valuvat Eulerin menetelmällä laskettaessa hiljalleen ylöspäin, vaikka yhtälöiden tarkka ratkaisu on täysin jaksollinen.

Myös Runge–Kutta yleistyy samalla tavalla useamman yhtälön tapaukseen, ja peittoa edelleen ylivoimaisesti Eulerin menetelmän laskentatarkkuudessa. Toki jos yhtälöitä on kaksi, niin Runge–Kutta-menetelmässä täytyy laskea joka askeleella neljän kulmakertoimen (47) sijaan kahdeksan.

Kaavojen lyhentämiseen on usein tapana käyttää vektorinotaatiota ratkaisulle, esimerkiksi jänikset ja sudet -esimerkissä kaksi populaatiofunktioita voisi mieltää vektoriarvoisena funktiona

$$\mathbf{y}(t) = \begin{bmatrix} R(t) \\ W(t) \end{bmatrix},$$

eli $\mathbf{y}(t)$ on ajasta t riippuva vektoriarvoinen funktio, jonka komponentit ovat populaatiokoot $R(t)$ ja $W(t)$. Tällöin populaatiomallin differentiaaliyhtälön (48) voi kirjoittaa myös vektorimuodossa

$$\mathbf{y}'(t) = h(\mathbf{y}(t)), \quad \text{missä} \quad \mathbf{y}'(t) = \begin{bmatrix} R'(t) \\ W'(t) \end{bmatrix},$$

ja h on vektoriarvoinen funktio, joka saadaan tässä tapauksessa kaavasta

$$h(\mathbf{v}) = \begin{bmatrix} k_1 v_1 - e_1 v_1 v_2 \\ e_2 v_1 v_2 - k_2 v_2 \end{bmatrix}, \quad \text{missä} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Runge–Kutta-menetelmässä määritettävät yhtälöt voidaan myös ilmaista näppärämmin vektorimuodossa. Tutkitaan yleistä ensimmäisen asteen differentiaaliyhtälöryhmää

$$\mathbf{y}'(t) = h(t, \mathbf{y}(t)), \quad \mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix},$$

missä funktio h on sekä ajasta että vektorista $\mathbf{y}$ riippuva vektoriarvoinen funktio.

Tällöin Runge–Kutta-menetelmän kaavat (47) menevät vektorimuotoon

$$\begin{aligned} \mathbf{k}_1 &= h(t_n, \mathbf{y}_n) \\ \mathbf{k}_2 &= h\left(t_n + \frac{\Delta t}{2}, \mathbf{y}_n + \frac{\Delta t}{2} \cdot \mathbf{k}_1\right) \\ \mathbf{k}_3 &= h\left(t_n + \frac{\Delta t}{2}, \mathbf{y}_n + \frac{\Delta t}{2} \cdot \mathbf{k}_2\right) \\ \mathbf{k}_4 &= h(t_n + \Delta t, \mathbf{y}_n + \Delta t \cdot \mathbf{k}_3), \end{aligned}$$

missä $\mathbf{y}_n$ on ratkaisua approksimoiva vektori ajanhetkellä t_n . Lopulliset rekursio-kaavat ovat

$$t_{n+1} = t_n + \Delta t \quad \text{ja} \quad \mathbf{y}_{n+1} = \mathbf{y}_n + \frac{\Delta t}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4),$$

ja oletamme edelleen, että alkuhetkellä $t = t_0$ tunnemme ratkaisulle $\mathbf{y}$ alkuarvon $\mathbf{y}_0$, eli jokaisen yksittäisen ratkaisufunktion arvon ajanhetkellä t_0 .

Erityinen hyöty vektorinotaatiosta on, että yllä oleva kaava ei muutu, vaikka tuntemattomia funktioita olisi tuhat kappaletta kahden sijaan. Toki ratkaisuvektorissa $\mathbf{y}_n$ olisi tällöin tuhat komponenttia. Vektorilaskenta yksinkertaistaa myös menetelmien toteutusta tietokoneella. Esimerkiksi Python-ohjelmointikielen kirjasto `numpy` hoitaa monet vektori- ja matriisipohjaiset laskut kätevästi.

9.4 Korkeamman asteen yhtälöt numeerisesti

Tähän mennessä olemme käsitelleet ainoastaan ensimmäisen asteen differentiaaliyhtälöiden (ja -ryhmien) ratkaisemista numeerisesti, mutta osoittautuu, että korkeamman asteen yhtälötkään eivät tuota sen suurempia vaikeuksia. Tutkitaan esimerkiksi yleistä toisen asteen differentiaaliyhtälöä

$$y''(x) = h(x, y(x), y'(x)), \quad (49)$$

eli yhtälöä, jossa toisen asteen derivaatta y'' riippuu annetun funktion h määräämällä tavalla muuttujasta x , funktiosta y , ja derivaatasta y' .

Näppärä temppu muodostaa tälle yhtälölle numeerinen ratkaisu on nimetä uudelleen ensimmäisen asteen derivaatta funktioksi $z(x) = y'(x)$. Tämä aluksi varsin turhalta vaikuttava toimenpide muuttaa kuitenkin toisen asteen yhtälön (49) ensimmäisen asteen yhtälöryhmäksi

$$\begin{cases} y'(x) = z(x) \\ z'(x) = h(x, y(x), z(x)) \end{cases}.$$

Tällaisia yhtälöryhmiä opimme ratkaisemaan numeerisesti juuri edellisessä osiossa. Esimerkiksi jos asetetaan kohdassa $x = 0$ alkuarvot $y(0) = y_0$ ja $y'(0) = z(0) = z_0$,

niin Eulerin menetelmän rekursiokaavat saavat muodon

$$\begin{cases} y_{n+1} = y_n + \Delta x \cdot z_n \\ z_{n+1} = z_n + \Delta x \cdot h(x_n, y_n, z_n) \end{cases}.$$

Ratkaisuapproksimaation kuvaajaa piirtäessä voidaan apufunktion arvot z_n unohtaa, koska kuvaajan hahmottamiseen tarvitaan ainoastaan pisteet (x_n, y_n) . Lukuja z_n tarvittiin siis vain pitämään muistissa arvioita funktion y derivaatalle pisteissä x_n . Eulerin menetelmän soveltamisessa tähän tilanteeseen on siis lopulta kyse vain siitä, että ratkaisufunktion y ja sen derivaatan y' arvoja approksimoidaan joka askeleella lineaarisesti. Arvioita saadaan tarkennettua jälleen pienentämällä askelpituutta, ja toki vaihtamalla tehokkaampaan Runge–Kutta-menetelmään.

Tässä esitetty temppu toimii samalla tavalla myös korkeamman asteen yhtälöihin, kunhan lisätään vain riittävän monta apufunktiota. Esimerkiksi neljännen asteen yhtälö $y'''' = y$ voidaan kirjoittaa apufunktioiden avulla neljänä yhtälönä

$$y' = z, \quad z' = u, \quad u' = v, \quad v' = y.$$

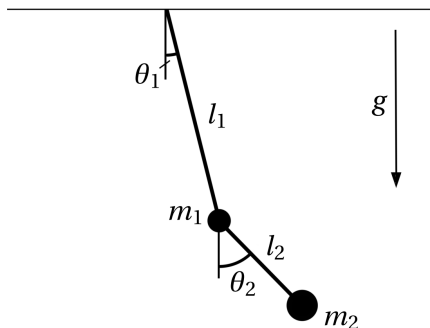
Lisäksi sama idea tepsii, vaikka tuntemattomia funktioita olisi useampia. Lukija voi nimittäin halutessaan tulkita yhtälön (49) myös vektorimuodossa, eli ajatella ratkaisufunktiota y , sen derivaattoja ja funktiota h vektoriarvoisena. Tällöin yhtälö (49) esittää usean tuntemattoman funktion differentiaaliyhtälöryhmää.

Toisen asteen differentiaaliyhtälöryhmiä esiintyy erityisesti paljon fysiikassa, esimerkiksi kitkattoman kaksoisheilurin liikettä kuvaavat (varsin monimutkaiset) yhtälöt

$$\theta_1'' = \frac{-g(2m_1 + m_2) \sin \theta_1 - m_2 g \sin(\theta_1 - 2\theta_2) - 2 \sin(\theta_1 - \theta_2) m_2 (\theta_2'^2 l_2 + \theta_1'^2 l_1 \cos(\theta_1 - \theta_2))}{l_1 (2m_1 + m_2 - m_2 \cos(2\theta_1 - 2\theta_2))}$$

$$\theta_2'' = \frac{2 \sin(\theta_1 - \theta_2) (\theta_1'^2 l_1 (m_1 + m_2) + g(m_1 + m_2) \cos \theta_1 + \theta_2'^2 l_2 m_2 \cos(\theta_1 - \theta_2))}{l_2 (2m_1 + m_2 - m_2 \cos(2\theta_1 - 2\theta_2))}.$$

Apufunktioiden $\omega_1(t) = \theta_1'(t)$ ja $\omega_2(t) = \theta_2'(t)$ avulla tällainen kahden funktion toisen asteen yhtälöryhmä palautuu neljän funktion ensimmäisen asteen yhtälöryhmäksi, joka voidaan ratkaista numeerisilla menetelmillä tavalliseen tapaan.



Kuva 42. Kaksoisheilurissa kaksi kappaletta (massat m_1 ja m_2) heiluvat kahden tangon (pituudet l_1 ja l_2) varassa painovoiman vaikutuksen alla.

Harjoitustehtäviä

Tehtävä 74. Ratkaise käsin laskien alkuarvo-ongelmasta

$$y' = xy, \quad y(1) = 1$$

likiarvot luvuille $y(2)$ ja $y(3)$. Käytä Eulerin menetelmää askelpituudella $\Delta x = 1$. Vertaa tarkkaan ratkaisuun $y(x) = e^{\frac{1}{2}x^2 - \frac{1}{2}}$.

Vihje s. 163 →

Toinen vihje s. 180 →

Tehtävä 75. Selvitä käsin laskien Runge–Kutta-menetelmällä likiarvo luvulle $y(2)$ tehtävän 74 alkuarvo-ongelmassa

$$y' = xy, \quad y(1) = 1.$$

Käytä edelleen askelpituutta $\Delta x = 1$. Kuinka suuri virhe on nyt?

Vihje s. 164 →

Toinen vihje s. 180 →

Huom! Seuraavat tehtävät saattavat vaatia laskinta tai pientä ohjelmointia haluamallasi ohjelmointikielellä. Myös taulukkolaskenta on hyvä vaihtoehto, sillä sen avulla välitulokset näkyvät selkeästi ja kuvaajien piirto on helppoa.

Tehtävä 76. Etsi Eulerin menetelmällä Neperin luvulle e likiarvo tutkimalla alkuarvo-ongelmaa $y' = y$ ja $y(0) = 1$.

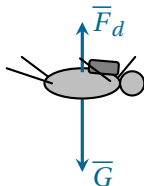
a) Arvioi lauseketta $y(1)$ käyttämällä askelpituutta $\Delta x = 0,1$.

b) Minkä arvion saat, kun asetat yleisen askelpituuden $1/N$?

Vihje s. 164 →

Toinen vihje s. 180 →

Tehtävä 77. Mallinnetaan laskuvarjohypyn alkua mallilla, jossa hyppääjään massa on m , hyppääjään vaikuttaa alaspäin painovoima $G = mg$ ja ylöspäin ilmanvastus $F_d = qv^2$, missä v on hyppääjän nopeus ja q vakio.



Mallinna hyppääjän nopeuden kehitystä Eulerin menetelmällä hypyn ensimmäisen 20 sekunnin aikana, kun alkunopeus on $v(0) = \frac{m}{s}$, $m = 75\text{kg}$, $k = 0,26\frac{\text{kg}}{\text{s}}$, ja $g = 9,81\frac{\text{m}}{\text{s}^2}$. Käytä askelpituutta $\Delta t = 1\text{s}$.

Vihje s. 164 →

Toinen vihje s. 181 →

Tehtävä 78. Tutkitaan differentiaaliyhtälöä $y' = -1000y$, jolle löytyy ratkaisu $y(x) = e^{-1000x}$ alkuarvolla $y(0) = 1$. Tutki, mitä tapahtuu Eulerin menetelmää soveltaessa ensimmäisellä askeleella, kun askelpituus on $\Delta x = 0,1$. Miksi tämä on ongelma?

Vaihtoehtoinen tapa ratkaista yhtälö numeerisesti on soveltaa **käänteistä Eulerin menetelmää**, jossa seuraava approksimaatiopiste y_{n+1} ratkaistaan implisiittisestä kaavasta

$$y_{n+1} = y_n + h(x_{n+1}, y_{n+1})\Delta x.$$

Käänteinen Eulerin menetelmä siis perustuu funktion y derivaatan ennustamiseen seuraavassa pisteessä x_{n+1} edellisen pisteen x_n sijaan. Tämä menetelmä on erityisesti kannattavaa, kun funktion y derivaatassa tapahtuu suuria muutoksia pienellä välillä.

Laske käänteisellä Eulerin menetelmällä approksimaatio yhtälön $y' = -1000y$ ratkaisulle välillä $[0,1]$ alkuarvolla $y(0) = 1$ ja askelpituudella $\Delta x = 0,1$. Vertaa ratkaisua ja approksimaatiota pisteessä $x = 1$.

Vihje s. 165 →

Toinen vihje s. 181 →

Tehtävä 79. Tutki alkuarvo-ongelman $y' = y + \sin(x) + 2$ ja $y(0) = 1$ ratkaisun arvioimista

- Eulerin menetelmällä
- Runge–Kutta-menetelmällä.
- Vertaa menetelmien virhettä todelliseen ratkaisun arvoon pisteessä $x = 4$ askelpituudella $\Delta x = 0,2$.

Vihje s. 165 →

Toinen vihje s. 181 →

Tehtävä 80. Mallinnetaan zombieepidemiaa differentiaaliyhtälöryhmällä

$$\begin{cases} H'(t) = -\beta H(t)Z(t) \\ Z'(t) = \beta H(t)Z(t) - \alpha(1 + \gamma Z(t))H(t)Z(t) \end{cases}.$$

Tässä $H(t)$ kuvastaa ihmisten lukumäärää ja $Z(t)$ kuvastaa zombien lukumäärää. Parametrit $\alpha, \beta, \gamma > 0$ vaikuttavat tilanteen kehitykseen. Pohdi ensin, mitä ilmiöitä mallissa esiintyvät lausekkeet oikeastaan esittävät zombi-invaasion kannalta. Tutki sitten ratkaisua Eulerin menetelmällä alkuarvoilla $H(0) = 400$, $Z(0) = 10$, ja parametreilla $\alpha = 0,0006$, $\beta = 0,002$, ja $\gamma = 0,1$ aikavälillä $[0, 60]$. Mikä on loppuratkaisu?

Vihje s. 165 →

Toinen vihje s. 181 →

Liitteet

Liite A: Osittaisintegrointi

Osittaisintegrointikaava on tulon derivointikaavan integraaliversio. Se kertoo, että jatkuvasti derivoituville funktioille f ja g pätee

$$\int f'(x) \cdot g(x) \, dx = f(x)g(x) - \int f(x) \cdot g'(x) \, dx.$$

Osittaisintegrointikaavan todistus on suoraviivainen: kun tulon derivointikaavaa

$$\frac{d}{dx} [f(x) \cdot g(x)] = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

integroidaan puolittain muuttujan x suhteen, saadaan

$$f(x) \cdot g(x) + C = \int f'(x) \cdot g(x) \, dx + \int f(x) \cdot g'(x) \, dx,$$

missä C on integrointivakio. Tästä voidaan ratkaista

$$\int f'(x) \cdot g(x) \, dx = f(x) \cdot g(x) - \int f(x) \cdot g'(x) \, dx + C,$$

ja integrointivakio voidaan lopuksi yhdistää sen viereisen integraalin integrointivakioon. Lauseessa esiintyvät integraalit ovat hyvin määriteltäviä, mikäli f' ja g' ovat jatkuvia.

Esimerkiksi integraali $\int x \cdot \cos x \, dx$ voitaisiin laskea osittaisintegroinnin avulla tulkinnalla $f(x) = \sin x$ ja $g(x) = x$, jolloin $f'(x) = \cos x$ ja $g'(x) = 1$. Saadaan

$$\int \underbrace{\cos(x)}_{f'(x)} \cdot \underbrace{x}_{g(x)} \, dx = \underbrace{\sin(x)}_{f(x)} \cdot \underbrace{x}_{g(x)} - \int \underbrace{\sin(x)}_{f(x)} \cdot \underbrace{1}_{g'(x)} \, dx = \sin(x) \cdot x + \cos x + C.$$

Osittaisintegrointikaava määrätyille integraaleille on luontevasti

$$\int_a^b f'(x) \cdot g(x) \, dx = \Big/ \int_a^b f(x)g(x) - \int_a^b f(x) \cdot g'(x) \, dx,$$

sillä summassa sijoituksen voi hyvin tehdä termi kerrallaan.

Liite B: Muuttujanvaihto integraalissa

Yhdistetyn funktion derivointikaavasta on olemassa vastaava kaava integraaleille, ja sitä kutsutaan integraalin **muuttujanvaihtokaavaksi**. Kaava kertoo, että jos funktio g on suljetulla välillä $[a, b]$ jatkuva ja avoimella välillä $]a, b[$ jatkuvasti derivoituva, ja lisäksi funktio f on suljetulla välillä $[g(a), g(b)]$ jatkuva, niin

$$\int_a^b f(g(t))g'(t) dt = \int_{g(a)}^{g(b)} f(x) dx.$$

Tässä jälkimmäisessä integraalissa käytetään muuttujaa x muuttujan t sijaan selkeyden vuoksi, sillä muuttujat t ja x saattavat kuulua eri väleille.

Käydään ensin läpi kaavan todistus. Olkoon funktio $F(t)$ funktion $f(t)$ eräs integraalifunktio, eli $F'(t) = f(t)$. Tällöin yhdistetyn funktion derivointikaava antaa

$$\frac{d}{dt}F(g(t)) = f(g(t))g'(t).$$

Siispä funktio $F(g(t))$ on funktion $f(g(t))g'(t)$ eräs integraalifunktio, joten haluamme kaava seuraa laskusta

$$\begin{aligned} \int_a^b f(g(t))g'(t) dt &= \int_a^b F(g(t)) \\ &= F(g(b)) - F(g(a)) \\ &= \int_{g(a)}^{g(b)} F(x) \\ &= \int_{g(a)}^{g(b)} f(x) dx. \end{aligned}$$

Muuttujanvaihtokaava esiintyy tilanteesta riippuen myös hieman eri muodoissa. Mikäli funktiolla g on käänteisfunktio g^{-1} , voidaan nimetä $c = g(a)$ ja $d = g(b)$ ja esittää kaava vaihtoehtoisessa muodossa

$$\int_c^d f(x) dx = \int_{g^{-1}(c)}^{g^{-1}(d)} f(g(t))g'(t) dt. \quad (50)$$

Tätä kaavaa on usein tapana ajatella sijoituksena $x = g(t)$, eli muuttujan x paikalle sijoitetaan lauseke $g(t)$. Tällöin differentiaalinen dx paikalle vaihdetaan kaavan mukaan $g'(t) dt$, minkä voi muistaa laskemalla

$$\frac{dx}{dt} = g'(t), \quad \text{mistä (epätäsmällisesti) kertomalla saadaan} \quad dx = g'(t) dt.$$

Sopiva sijoitus voi toisinaan tehdä vaikeasta integraalista helpomman. Lasketaan esimerkkinä määrätty integraali

$$\int_{-1}^1 \sqrt{1-x^2} dx,$$

joka voidaan tulkita origokeskeisen 1-säteisen puoliympyrän pinta-alana, sillä lauseke $\sqrt{1-x^2}$ esittää ympyrän korkeutta kohdassa x . Tämä integraali voidaan laskea analyyttisesti sijoituksella $x = g(t) = \sin(t)$, missä muuttuja t liikkuu välillä $[-\pi/2, \pi/2]$, sillä $\sin(-\pi/2) = -1$ ja $\sin(\pi/2) = 1$. Funktio $\sin(t)$ on myös aidosti kasvava välillä $[-\pi/2, \pi/2]$, ja täten sillä on käänteisfunktio tällä välillä. Nyt $g'(t) = \cos t$, joten muuttujanvaihtokaavan (50) perusteella saadaan

$$\begin{aligned} \int_{-1}^1 \sqrt{1-x^2} dx &= \int_{-\pi/2}^{\pi/2} \sqrt{1-\sin^2(t)} \cos(t) dt \\ &= \int_{-\pi/2}^{\pi/2} \sqrt{\cos^2(t)} \cos(t) dt \\ &= \int_{-\pi/2}^{\pi/2} \cos^2(t) dt \\ &= \int_{-\pi/2}^{\pi/2} \frac{\cos(2t) + 1}{2} dt \\ &= \int_{-\pi/2}^{\pi/2} \left(\frac{\sin(2t)}{4} + \frac{t}{2} \right) \\ &= 0 + \frac{\pi}{4} - \left(0 - \frac{\pi}{4} \right) \\ &= \pi/2, \end{aligned}$$

missä hyödynsimme kosinfunktion ei-negatiivisuutta välillä $[-\pi/2, \pi/2]$ neliöjuuren poistamisessa ja kaavaa $\cos^2(t) = \frac{1}{2}(\cos(2t) + 1)$ integroitavan lausekkeen sieventämisessä. Siis puoliympyrän pinta-ala todellakin on $\pi/2$.

Muuttujanvaihtokaava toimii samalla tavalla myös määräämättömille integraaleille, eli

$$\int f(g(t))g'(t) dt = \int f(x) dx.$$

Liite C: Integraalilaskennan yleistetty väliarvolause

Integraalilaskennan yleistetty väliarvolause sanoo, että mikäli f on jatkuva ja g integroituva funktio välillä $x \in [a, b]$, ja funktio g ei vaihda merkkiään tällä välillä (eli joko $g(x) \geq 0$ tai $g(x) \leq 0$), pätee

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

missä t on jokin luku välillä $]a, b[$.

Todistetaan lause erikoistapauksessa, jossa g on jatkuva funktio ja koko välillä $[a, b]$ pätee $g(x) > 0$. Jos funktion f suurin arvo välillä $[a, b]$ on M ja pienin arvo luku m , voidaan arvioida seuraavasti:

$$m \int_a^b g(x) \, dx = \int_a^b m \cdot g(x) \, dx \leq \int_a^b f(x)g(x) \, dx \leq \int_a^b M \cdot g(x) \, dx = M \int_a^b g(x) \, dx.$$

Koska määrätty integraali $\int_a^b g(x) \, dx$ on tässä tapauksessa positiivinen vakio, voidaan jakaa sillä ja saadaan

$$m \leq \frac{\int_a^b f(x)g(x) \, dx}{\int_a^b g(x) \, dx} \leq M.$$

Koska jatkuva funktio f saa kaikki arvot suurimman arvonsa M ja pienimmän arvonsa m välillä, jollakin muuttujan arvolla $t \in]a, b[$ pätee

$$f(t) = \frac{\int_a^b f(x)g(x) \, dx}{\int_a^b g(x) \, dx}, \quad \text{eli on todistettu} \quad f(t) \cdot \int_a^b g(x) \, dx = \int_a^b f(x)g(x) \, dx.$$

Liite D: Taylorin polynomin virhekaavojen todistukset

Todistamme nyt Taylorin polynomin virhekaavat aloittaen ensimmäisestä virhekaavasta. Tarvitsemme pohjatietoina kaksi tulosta, jotka eivät kirjoitushetkellä kuuluneet lukion oppimäärään: Osittaisintegrointi (Liite A) ja Integraalilaskennan yleistetty väliarvolause (Liite C).

Näiden pohjatietojen avulla saamme todistettua itse virhekaavan.

Väite: Taylorin lause. Olkoon f funktio, joka on $n + 1$ kertaa jatkuvasti derivoituva reaalilukujen a ja x muodostamalla avoimella välillä, ja jonka n . derivaatta $f^{(n)}$ on jatkuva lukujen a ja x välisellä suljetulla välillä.

Tällöin funktiolle f kohdassa $x = a$ kehitetyn Taylorin polynomin $T_n(x)$ virhe kohdassa x on

$$R_n(x) = \int_a^x \frac{f^{(n+1)}(s)}{n!} (x-s)^n ds = \frac{f^{(n+1)}}{(n+1)!} (x-a)^{n+1}.$$

Todistus. Ensinnäkin määrätyn integraalin perusominaisuuksista seuraa, että

$$\int_a^x f'(s) ds = \int_a^x f(s) = f(x) - f(a), \quad \text{eli} \quad f(x) = f(a) + \int_a^x f'(s) ds.$$

Muokataan nyt tätä viimeistä integraalia integroimalla osittain niin, että funktiota $f'(s)$ derivoidaan, ja toinen funktio on $g(s) = s - x$, jolloin $g'(s) = 1$.

$$\begin{aligned} \int_a^x f'(s) ds &= \int_a^x \underbrace{1}_{g'(s)} \cdot f'(s) ds = \int_a^x \underbrace{(s-x)}_{g(s)} f'(s) - \int_a^x \underbrace{(s-x)}_{g(s)} f''(s) ds \\ &= (x-x)f'(x) - (a-x)f'(a) - \int_a^x (s-x)f''(s) ds \\ &= (x-a)f'(a) + \int_a^x (x-s)f''(s) ds. \end{aligned}$$

Kun tämä sijoitetaan aiempaan funktion $f(x)$ lausekkeeseen, saadaan lupaavasti

$$f(x) = f(a) + f'(a)(x-a) + \int_a^x (x-s)f''(s) ds.$$

Seuraavaksi osittaisintegroidaan taas jäljellä olevaa integraalia. Tällä kertaa derivoidaan funktiota $f''(s)$ ja integroidaan funktiota $(x-s)$, jonka integraali on $-\frac{1}{2}(x-s)^2$. Saadaan

$$\begin{aligned} \int_a^x (x-s)f''(s) ds &= \int_a^x -\frac{1}{2}(x-s)^2 f''(s) - \int_a^x -\frac{1}{2}(x-s)^2 f'''(s) ds \\ &= \left[-\frac{1}{2}(x-x)^2 f''(x)\right] - \left[-\frac{1}{2}(x-a)^2 f''(a)\right] + \int_a^x \frac{1}{2}(x-s)^2 f'''(s) ds \\ &= \frac{f''(a)}{2}(x-a)^2 + \int_a^x \frac{f'''(s)}{2}(x-s)^2 ds. \end{aligned}$$

Nyt funktiolle $f(x)$ on kasassa lauseke

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \int_a^x \frac{f'''(s)}{2}(x-s)^2 ds.$$

Taylorin polynomi alkaa hahmottua! Kun samaa prosessia jatketaan, n osittaisintegroinnin jälkeen ollaan muodossa

$$f(x) = \underbrace{f(a) + f'(a)(x-a) + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n}_{T_n(x)} + \underbrace{\int_a^x \frac{f^{(n+1)}(s)}{n!}(x-s)^n ds}_{R_n(x)}.$$

Näin Taylorin polynomin virheelle kohdassa x on siis johdettu integraalimuotoinen kaava. Virhekaavan toiseen muotoon päästään soveltamalla integraalilaskennan yleistettyä väliarvolausetta ja integroimalla jäljelle jäänyt polynomi:

$$\begin{aligned} R_n(x) &= \int_a^x \frac{f^{(n+1)}(s)}{n!}(x-s)^n ds = f^{(n+1)}(t) \int_a^x \frac{1}{n!}(x-s)^n ds \\ &= f^{(n+1)}(t) \cdot \int_a^x \frac{-1}{(n+1)!}(x-s)^{n+1} = f^{(n+1)}(t) \cdot \frac{1}{(n+1)!}(x-a)^{n+1}. \quad \square \end{aligned}$$

Toisen virhekaavan todistus

Esitämme seuraavaksi Taylorin toisen virhekaavan todistuksen idean. Haluamme siis osoittaa, että

$$f(x) = T_n(x) + r_n(x)(x-a)^n,$$

missä $r_n(x)$ on funktio, jolle $\lim_{x \rightarrow a} r_n(x) = 0$.

Sovelletaan tässä vaiheessa edellisessä todistuksessa johdettua integraalimuotoista esitystä asteen $n-1$ Taylorin polynomin virheelle, jonka mukaan

$$f(x) = \underbrace{f(a) + f'(a)(x-a) + \dots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}}_{T_{n-1}(x)} + \underbrace{\int_a^x \frac{f^{(n)}(s)}{(n-1)!}(x-s)^{n-1} ds}_{R_{n-1}(x)}.$$

Tuotetaan oikealle puolelle asteen n Taylorin polynomi T_n lisäämällä ja vähentämällä puuttuva termi $\frac{f^{(n)}(a)}{n!}(x-a)^n$, josta saadaan

$$f(x) = T_n(x) + \int_a^x \frac{f^{(n)}(s)}{(n-1)!}(x-s)^{n-1} ds - \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Haluaisimme muuttaa viimeisen termin integraalimuotoon, joten lasketaan, että

$$\frac{f^{(n)}(a)}{n!}(x-a)^n = \frac{f^{(n)}(a)}{n!} \int_a^x n(x-s)^{n-1} ds = \int_a^x \frac{f^{(n)}(a)}{(n-1)!}(x-s)^{n-1} ds.$$

Yhdistämällä tämä edelliseen kaavaan saadaan

$$f(x) = T_n(x) + \underbrace{\int_a^x \frac{f^{(n)}(s) - f^{(n)}(a)}{(n-1)!} (x-s)^{n-1} ds}_{R_n(x)}.$$

Tästä uudesta muodosta virhetermille $R_n(x)$ voidaan nyt löytää arvio, joka perustuu asteen n derivaatan $f^{(n)}$ jatkuvuuteen. Aloitetaan määrittelemällä lauseke $M_x = \max_{a \leq s \leq x} |f^{(n)}(s) - f^{(n)}(a)|$, joka vähenee kohti nollaa, kun x lähestyy pistettä a funktion $f^{(n)}$ jatkuvuuden perusteella. Saadaan nyt arvio

$$\begin{aligned} |R_n(x)| &= \left| \int_a^x \frac{f^{(n)}(s) - f^{(n)}(a)}{(n-1)!} (x-s)^{n-1} ds \right| \\ &\leq \int_a^x \frac{|f^{(n)}(s) - f^{(n)}(a)|}{(n-1)!} |x-s|^{n-1} ds \\ &\leq \frac{M_x}{n!} \int_a^x n|x-s|^{n-1} ds \\ &= \frac{M_x}{n!} |x-a|^n, \end{aligned}$$

missä viimeisessä välivaiheessa integraalin laskemiseen voidaan käsitellä tapaukset $x < a$ ja $x \geq a$ erikseen, jotta itseisarvomerkki ei häiritse. Tästä nähdään lopulta, että

$$|r_n(x)| = \frac{|R_n(x)|}{|x-a|^n} \leq \frac{M_x}{n!} \rightarrow 0, \quad \text{kun } x \rightarrow a.$$

Siis toisen virhekaavan väittämä on todistettu. Erityisesti todistuksesta voidaan huomata, että virhearviossa esiintyvän tekijän $r_n(x)$ suppenemisvauhti nollaan riippuu pääsääntöisesti siitä, kuinka nopeasti lauseke $f^{(n)}(x)$ suppenee kohti arvoa $f^{(n)}(a)$, eli siitä kuinka vahvasti asteen n derivaatta $f^{(n)}$ on jatkuva pisteessä a . Jatkuvuuden määritelmää käsitellään tarkemmin Luvussa 3.

Liite E: Newtonin menetelmän suppeneminen

Todistamme tässä sivulla 5.2 esitellyt tulokset Newtonin menetelmän suppenemisestä.

Tulos 1. Oletetaan, että funktiolla f on nollakohta a , funktio f on derivoituva jossakin kohdan $x = a$ ympäristössä, derivaattafunktio f' on tässä ympäristössä jatkuva, ja $f'(a) \neq 0$. Väitämme, että kun alkuarvo x_0 valitaan riittävän läheltä nollakohtaa $x = a$, Newtonin menetelmä suppenee tähän nollakohtaan.

Menetelmän suppeneminen perustuu siihen, että jos funktio f on jatkuvasti derivoituva, niin nollakohdan lähettyvillä sen kuvaaja kulkee hyvin lähellä Newtonin menetelmässä muodostettua tangenttisuoraa ja täten menetelmä antaa hyviä arvioita. Tehdään täsmällinen tarkastelu.

Olkoon x_0 nyt jokin piste, joka on valittu läheltä funktion nollakohtaa a . Oletusten perusteella siis $f(a) = 0$ ja $f'(a) \neq 0$. Muodostetaan funktiolle pisteen a suhteen ensimmäisen asteen Taylorin kehitemä, ja sijoitetaan siihen piste x_0 . Tästä saadaan

$$f(x_0) = f'(a)(x_0 - a) + R_1(x_0).$$

Tässä vakiotermi hävisi, koska $f(a) = 0$, ja virhetermi $R_1(x_0)$ voidaan tarvittaessa arvioida pieneksi. Nyt seuraava iteraatiopiste x_1 saadaan kaavasta

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{f'(a)(x_0 - a) + R_1(x_0)}{f'(x_0)}.$$

Lopullinen tavoitteemme on osoittaa, että erotus $x_1 - a$ on pieni, joten vähennetään puolittain a ja muokataan yhtälöä seuraavasti:

$$\begin{aligned} x_1 - a &= x_0 - a - \frac{f'(a)(x_0 - a) + R_1(x_0)}{f'(x_0)} \\ &= \frac{f'(x_0)(x_0 - a) - f'(a)(x_0 - a) + R_1(x_0)}{f'(x_0)} \\ &= \frac{(f'(x_0) - f'(a))(x_0 - a)}{f'(x_0)} + \frac{R_1(x_0)}{f'(x_0)}. \end{aligned}$$

Tarkistellaan nyt näitä kahta virhetermiä erikseen. Koska derivaatta f' on jatkuva, voimme raja-arvon määritelmän (Luku 3) perusteella valita pisteen x_0 niin läheltä pistettä a , että lukujen $f'(x_0)$ ja $f'(a)$ etäisyys on pienempi kuin haluamamme positiivinen luku. Koska $f'(a) > 0$, voimme valita täksi luvuksi $\frac{1}{4} |f'(a)|$. Sopivalla x_0 pätee siis

$$|f'(x_0) - f'(a)| \leq \frac{1}{4} |f'(a)|, \quad \text{josta seuraa} \quad |f'(x_0)| \geq \frac{1}{2} |f'(a)|.$$

Arvioimme nyt ensimmäisen virhetermin itseisarvoa seuraavasti

$$\frac{|(f'(x_0) - f'(a))(x_0 - a)|}{|f'(x_0)|} \leq \frac{\frac{1}{4}|f'(a)||x_0 - a|}{\frac{1}{2}|f'(a)|} = \frac{1}{2}|x_0 - a|.$$

Jälkimmäiselle termille voidaan hyödyntää lisäksi Taylorin polynomin toista virhekaavaa, josta saadaan arvio

$$\frac{|R_1(x_0)|}{|f'(x_0)|} = \frac{|r_1(x_0)|}{|f'(x_0)|}|x_0 - a| \leq \frac{2|r_1(x_0)|}{|f'(a)|}|x_0 - a|.$$

Tässä kerroin $2|r_1(x_0)|/|f'(a)|$ voidaan valita mielivaltaisen pieneksi, sillä virhekaavan perusteella $\lim_{x_0 \rightarrow a} r_1(x_0) = 0$. Valitaan x_0 niin läheltä pistettä a , että kerroin on pienempää kuin $1/4$, jolloin yhdistämällä edelliset kaksi arviota saadaan lopulta, että

$$|x_1 - a| \leq \frac{1}{2}|x_0 - a| + \frac{1}{4}|x_0 - a| = \frac{3}{4}|x_0 - a|.$$

Tämä arvio riittää meille, sillä sen mukaan Newtonin menetelmässä seuraava piste on ykköstä pienemmän vakiokertoimen verran lähempänä nollakohtaa kuin edellinen. Menetelmä suppenee siis ainakin eksponentiaalisesti. Todellisuudessa suppenemisvauhti on hieman nopeampi, sillä yllä olevat arviot paranevat ennestään pisteiden lähentyessä nollakohtaa a , joten vakiokertoimen $3/4$ paikalle saadaan mielivaltaisen pieni kerroin, kun mennään riittävän lähelle pistettä a .

Todistetaan seuraavaksi vahvempi suppenemisehto.

Neliöllinen suppeneminen. Oletetaan lisäksi, että funktion f toinen derivaatta f'' on olemassa pisteen a lähetyvillä, ja että sille pätee siellä yläarvio $|f''(x)| \leq M$ jollain vakiolla M . Tällöin edellistä kahta virhetermiarviota voidaan parantaa ennestään.

Aloitetaan ensin antamalla parempi arvio erotukselle $f'(x_0) - f'(a)$. Väliarvolauseen perusteella löytyy jokin piste ξ pisteiden a ja x_0 välissä siten, että $f'(x_0) - f'(a) = f''(\xi)(x_0 - a)$. Siispä

$$|f'(x_0) - f'(a)| = |f''(\xi)||x_0 - a| \leq M|x_0 - a|.$$

Käyttämällä tätä hyödyksi edellisen laskun ensimmäisen virhetermin arvioissa saadaan parempi arvio

$$\frac{|f'(x_0) - f'(a)|}{|f'(x_0)|}|x_0 - a| \leq \frac{2M}{|f'(a)|}|x_0 - a|^2.$$

Lisäksi Taylorin polynomin ensimmäisestä virhekaavasta saadaan toiselle termille parempi arvio

$$\frac{|R_1(x_0)|}{|f'(x_0)|} = \frac{|f''(\xi)||x_0 - a|^2}{2|f'(x_0)|} \leq \frac{M}{|f'(a)|}|x_0 - a|^2.$$

Yhdistämällä nämä kaksi arviota saadaan nyt lopullinen arvio

$$|x_1 - a| \leq \frac{2M}{|f'(a)|} |x_0 - a|^2 + \frac{M}{|f'(a)|} |x_0 - a|^2 = \frac{3M}{|f'(a)|} |x_0 - a|^2.$$

Tämä arvio on melko voimakas, sillä se kertoo, että Newtonin menetelmässä seuraavan iteraatiopisteen absoluuttinen virhe nollakohtaan a on aina verrannollinen edellisen virheen neliöön. Tällainen suppeneminen on huomattavasti nopeampaa kuin aiemmin saavutettu eksponentiaalinen suppenemisvauhti, ja peittoaa täten myös helposti esimerkiksi puolitusmenetelmän.

Liite F: Eulerin menetelmän virhearvio

Todistetaan sivulla 132 mainittu väite siitä, että Eulerin menetelmässä tapahtuva virhe aikavälillä $[0, T]$ ja askelpituudella Δx on verrannollinen lausekkeeseen Δx . Haluamme siis verrata menetelmän antamia lukuarvoja y_n funktion todellisiin arvoihin $y(x_n)$, missä $y(x)$ on yhtälömme $y' = h(x, y)$ ratkaisu alkuarvolla $y(0) = y_0$. Oletamme tässä, että funktio h ja sen derivaatat ovat kaikkialla hyvin määriteltyjä.

Muistetaan, että Eulerin menetelmässä pisteestä (x_n, y_n) muodostetaan seuraava piste (x_{n+1}, y_{n+1}) kaavoilla

$$x_{n+1} = x_n + \Delta x \quad \text{ja} \quad y_{n+1} = y_n + h(x_n, y_n) \Delta x.$$

Haluaisimme arvioida pisteen y_{n+1} virhettä ratkaisufunktion arvoon $y(x_{n+1})$. Tämä vaatii myös sen, että tunnemme jo tapahtuneen virheen edellisessä vaiheessa, eli luvun $V_n = y(x_n) - y_n$. Siis V_n ilmaisee joka askeleella n , kuinka kaukana ollaan ratkaisufunktion todellisesta arvosta pisteessä x_n .

Kirjoitetaan nyt funktiolle $y(x)$ ensimmäisen asteen Taylorin kehitelmä pisteen x_n suhteen. Tästä saadaan, että

$$y(x_{n+1}) = y(x_n) + y'(x_n)(x_{n+1} - x_n) + R_1(x_{n+1}) = y(x_n) + h(x_n, y(x_n)) \Delta x + R_1(x_{n+1}).$$

Otetaan tästä erotus pisteeseen y_{n+1} , ja saadaan rekursiokaavasta

$$\begin{aligned} V_{n+1} &= y(x_{n+1}) - y_{n+1} \\ &= y(x_n) - y_n + h(x_n, y(x_n)) \Delta x - h(x_n, y_n) \Delta x + R_1(x_{n+1}) \\ &= V_n + (h(x_n, y(x_n)) - h(x_n, y_n)) \Delta x + R_1(x_{n+1}). \end{aligned}$$

Arvioidaan nyt kahta viimeistä termiä. Ensimmäinen termi vaatii väliarvolauseen soveltamista funktion $h(x, y)$ osittaisderivaatalle y -suuntaan. Väliarvolauseen mukaan pisteiden y_n ja $y(x_n)$ välissä on jokin piste ξ , jolle

$$(h(x_n, y(x_n)) - h(x_n, y_n)) \Delta x = h_y(x_n, \xi) (y(x_n) - y_n) \Delta x = h_y(x_n, \xi) V_n \Delta x.$$

Jos M_1 on jokin yläraja osittaisderivaatalle $h_y(x, y)$, niin tämä lauseke on siis itseisarvoltaan enintään $M_1|V_n|\Delta x$. Tämä riittää meille. Arvioidaan vielä jälkimmäistä termiä Taylorin polynomin ensimmäisellä virhekaavalla, jonka mukaan

$$|R_1(x_{n+1})| = \left| \frac{y''(t)}{2} (x_{n+1} - x_n)^2 \right| \leq \frac{M_2}{2} (\Delta x)^2,$$

missä M_2 on yläraja ratkaisufunktion toiselle derivaatalle $y''(x)$ välillä $[0, T]$. Siis lopulta saadaan arvio

$$|V_{n+1}| \leq |V_n| + M_1|V_n|\Delta x + \frac{M_2}{2} (\Delta x)^2 = |V_n|(1 + M_1\Delta x) + \frac{M_2}{2} (\Delta x)^2.$$

Merkitään nyt $C = 1 + M_1\Delta x$. Alkuaskeleella pätee $y_0 = y(x_0)$, joten alussa virhettä ei ole, $V_0 = 0$. Ensimmäisen askeleen jälkeen virhe on siis enintään

$$|V_1| \leq \frac{M_2}{2} (\Delta x)^2.$$

Toisen askeleen jälkeen se on

$$|V_2| \leq C \frac{M_2}{2} (\Delta x)^2 + \frac{M_2}{2} (\Delta x)^2 = (C + 1) \frac{M_2}{2} (\Delta x)^2.$$

Tätä toistamalla huomataan yleinen arvio

$$|V_n| \leq (C^{n-1} + C^{n-2} + \dots + C + 1) \frac{M_2}{2} (\Delta x)^2 = \frac{C^n - 1}{C - 1} \frac{M_2}{2} (\Delta x)^2.$$

Toisaalta epäyhtälön $1 + x \leq e^x$ perusteella pätee $C^n = (1 + M_1\Delta x)^n \leq e^{nM_1\Delta x}$, joten saadaan lopulta.

$$|V_n| = (C^n - 1) \frac{M_2}{2M_1\Delta x} (\Delta x)^2 \leq (e^{nM_1\Delta x} - 1) \frac{M_2}{2M_1} \Delta x.$$

Välin $[0, T]$ sisällä tehdään askelia yhteensä $N = T/\Delta x$ kappaletta, joten $N\Delta x = T$. Täten lopullinen virhe välin päätepisteessä toteuttaa arvion

$$|V_N| \leq (e^{M_1 T} - 1) \frac{M_2}{2M_1} \Delta x = K\Delta x,$$

missä vakio $K = (e^{M_1 T} - 1) \frac{M_2}{2M_1}$ riippuu välin pituudesta, sekä ratkaisufunktion toisen derivaatan $y''(x)$ ja osittaisderivaatan $h_y(x, y)$ maksimiarvoista tällä välillä. Tämä on haluttu lopputulos.

Ensimmäiset vihjeet

Tehtävän 1 vihje. Määritelmän mukaan $\text{tarkka arvo} = \text{liikiarvo} + \text{virhe}$, joten virhe on tarkan arvon ja liikiarvon erotus. Virhe on tässä tapauksessa negatiivinen. Suhteellinen virhe on virheen ja tarkan arvon osamäärän itseisarvo.

← Tehtävä s. 19

Tehtävän 2 vihje. Laskinohjelmiston käyttämien liukulukujen tarkkuudesta riippuu, kuinka suuri eksponentti tarvitaan, ennen kuin $10^n + 1 - 10^n$ ei enää olekaan 1.

← Tehtävä s. 19

Tehtävän 3 vihje.

- a) Suurin luku on luonnollisesti $+9,9 \cdot 10^{+9} = 9\,900\,000\,000$. Huomionarvoista on, että toiseksi suurin on varsin paljon pienempi.
- b) Sovittiin, että nolasta poikkeava luku ei ala nolalla, joten pienimmän positiivisen luvun ensimmäinen numero on 1.
- c) Luvut 5 ja $1/5 = 0,2$ voidaan tallentaa tarkasti.
- d) Lukua $1/3 = 0,333\dots$ joudutaan pyöristämään.
- e) Koska käytössä on vain kaksi merkitsevää numeroa, ainakin toisen näistä kahdesta luvusta täytyy olla kaksinumeroinen.
- f) Ilmeisesti laskussa $100 + 2 + 2 + 2 + 2 + 2$ jokainen välitulos pyöristetään alaspäin, mikä selittää eron.
- g) Nolla kannattaa laskea erikseen, koska sen voi merkitä monella tavalla (kaikki eksponentit ja molemmat etumerkit käyvät). Nolasta poikkeavat luvut voi merkitä vain yhdellä tavalla, koska sovittiin, että niissä mantissa ei ole nolalla. Yhteensä pitäisi löytyä 3421 eri lukua.
- h) Kummallakin välillä on yhtä monta lukua, ero on vain eksponentissa.

← Tehtävä s. 19

Tehtävän 4 vihje.

- a) Voi olla kätevää pilkkoa luvut kakkosen potensseiksi. Esimerkiksi $7 = 4 + 2 + 1$, joten se on binäärilukuna 111, ja vastaavasti $5 = 4 + 1 = 101_2$.
- b) Jokainen paikka vastaa kakkosen potenssia, joten $1110_2 = 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$.
- c) Suurin lukua 21 pienempi kakkosen potenssi on $16 = 2^4$, joten binääriluvusta tulee viisinumeroinen.
- d) Voi olla aluksi helpointa muuttaa luvut kymmenjärjestelmään, laskea siellä ja muuttaa takaisin.
- e) Binääriluvuissa on vain kaksi yksinnumeroista epänegatiivista lukua, nimittäin 0 ja 1. Kumpaankin tauluun tulee siis vain $2 \times 2 = 4$ laskua.
- f) Allekkain laskeminen toimii samoin kuin kymmenjärjestelmässä. Kannattaa huomioida, että laskut $1 + 1 = 10$ ja $1 + 1 + 1 = 11$ tuottavat muistinumeron. Vastaukseksi pitäisi tulla 1011101.
- g) Vastaukseksi pitäisi tulla 1000001.
- h) Binääriluvun pilkun jälkeiset ensimmäiset paikat kuvaavat puolikkaita, neljäsosia ja kahdeksasosia. Voi olla helpointa käyttää välivaiheessa murtolukuja. Vastaukseksi pitäisi tulla 0,375.

← Tehtävä s. 20

Tehtävän 5 vihje. Luvun $-\frac{3}{8}$ voi pilkkoa muotoon $-\frac{3}{8} = -(\frac{1}{4} + \frac{1}{8})$. Miten tämän voisi esittää binäärilukuna?

← Tehtävä s. 20

Toinen vihje s. 166 →

Tehtävän 6 vihje. Kulmalla α on jokin oikea arvo, joka on pyöristyksen jälkeen päättynyt luvuksi $25,3^\circ$. Millä välillä kulman α oikea arvo voi olla?

← Tehtävä s. 20

Toinen vihje s. 166 →

Tehtävän 7 vihje. On tarkoitus laskea lausekkeiden

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

ja

$$\frac{f(1+h) - f(1-h)}{2h} = \frac{(1+h)^{1+h} - (1-h)^{1-h}}{2h}$$

liikiarvoja luvun h eri arvoilla. Pienillä luvun h arvoilla tulokset voivat riippua käytetystä ohjelmasta.

Derivaatan $f'(1)$ tarkkaa arvoa varten kannattaa hyödyntää kaavaa $x^x = e^{\ln(x^x)} = e^{x \ln x}$, joka auttaa derivaattafunktion $f'(x)$ määrittämisessä.

← Tehtävä s. 21

Tehtävän 8 vihje. Kohdan a) laskujen tulokset riippuvat käytetystä ohjelmasta. Todennäköinen lopputulos on, että suurilla luvun n arvoilla seurauksena on nollalla jako. Raja-arvo $\lim_{x \rightarrow \pi/3} L(x)$ on erään funktion derivaatta kohdassa $x = \frac{\pi}{3}$.

← Tehtävä s. 21

→ Toinen vihje s. 166

Tehtävän 9 vihje. Ongelma on siinä, että juurilauseke on liian lähellä ykköstä, jolloin vähennyslasku syö valtavasti tarkkuutta. Juuresta pitäisi siis päästä eroon jonkinlaisella sopivalla sievennyksellä. Muistikaava $(a+b)(a-b) = a^2 - b^2$ voisi tässä olla avuksi.

← Tehtävä s. 21

→ Toinen vihje s. 166

Tehtävän 10 vihje. Taylorin polynomin kertoimia varten tarvittaisiin funktion $f(x) = e^x$ arvo kohdassa $x = 0$, derivaatan arvo kohdassa $x = 0$ ja niin edelleen aina viidenteen derivaattaan asti. Onnekkaisesti eksponenttifunktion derivointi on varsin yksinkertaista. Saadut derivaatat tulisi sijoittaa Taylorin polynomin kaavaan

$$T_5(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \frac{f^{(4)}(a)}{4!}(x-a)^4 + \frac{f^{(5)}(a)}{5!}(x-a)^5,$$

missä $a = 0$.

← Tehtävä s. 32

→ Toinen vihje s. 166

Tehtävän 11 vihje. Aluksi olisi hyvä laskea funktion $f(x) = \sqrt{x}$ ensimmäinen ja toinen derivaatta $f'(x)$ ja $f''(x)$. Näiden avulla saadaan laskettua funktion arvo ja derivaattojen arvot kohdassa $x = 4$, eli luvut $f(4)$, $f'(4)$ ja $f''(4)$. Tuloksiksi pitäisi tulla 2 , $\frac{1}{4}$ ja $-\frac{1}{64}$.

← Tehtävä s. 32

→ Toinen vihje s. 167

Tehtävän 12 vihje. Funktion f kohdassa $x = a$ kehitetyn toisen asteen Taylorin polynomin kaava on

$$T_2(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2.$$

Nyt siis $a = 0$ ja $f(x) = \cos x$.

← Tehtävä s. 32

Toinen vihje s. 167 →

Tehtävän 13 vihje. Funktion f kohdassa $x = a$ kehitetyn toisen asteen Taylorin polynomin kaava on

$$T_2(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2.$$

Kaavaa tulisi siis käyttää kahdesti, ensin funktiolle $f(x) = \ln x + 1$, kun $a = 0$, ja sitten funktiolle $g(x) = \ln x$, kun $a = 1$.

← Tehtävä s. 32

Toinen vihje s. 167 →

Tehtävän 14 vihje. Taylorin polynomin virhekaavan mukaan virhe kohdassa x on

$$R_n(x) = \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{n+1}.$$

Tähän tulisi sijoittaa $n = 5$ ja oikeat arvot myös luvuille a ja x . Tarvitset myös funktion $f(x) = e^x$ kuudennen derivaatan. Luku t on tuntematon luku lukujen a ja x välissä.

← Tehtävä s. 32

Toinen vihje s. 167 →

Tehtävän 15 vihje. Voisi olla eduksi derivoida funktiot $f(x) = \sqrt{1+x}$ ja $g(x) = \sqrt[3]{1+x}$, ja tutkia, millä luvun x arvolla derivaatoiksi tulee $f'(x) = \frac{1}{2}$ ja $g'(x) = \frac{1}{3}$. Funktioille f ja g voisi kehittää Taylorin polynomit tämän kohdan ympäristössä.

← Tehtävä s. 32

Toinen vihje s. 167 →

Tehtävän 16 vihje. Kohta $x = 1$ kuuluu alueeseen $x > 0$, jolloin $f(x) = |x| = x$. Arvot $f(1)$, $f'(1)$, $f''(1)$ ja $f'''(1)$ ovat siis laskettavissa varsin suoraviivaisesti.

← Tehtävä s. 33

Toinen vihje s. 168 →

Tehtävän 17 vihje. Funktio $T_1(x) = x$ on sinifunktion $f(x) = \sin x$ ensimmäisen asteen Taylorin polynomi, joka on kehitetty kohdassa $x = 0$. Ensimmäisen asteen Taylorin polynomin virhekaava antaa kuitenkin virheen suuruudeksi

$$R_1(x) = \frac{f''(t)}{2!} \cdot (x-0)^2,$$

mikä on tehtävänantoa heikompi tulos. Mikä olisi sinin toisen asteen Taylorin polynomi $T_2(x)$?

← Tehtävä s. 33

Toinen vihje s. 168 →

Tehtävän 18 vihje. Hyvä lähtökohta olisi tarkastella Taylorin polynomin virhekaavaa asteen n Taylorin polynomille. Kaavassa esiintyy funktion $f(x) = \cos x$ asteen $n + 1$ derivaatta tuntemattomassa kohdassa $x = t$. Kosinin toistuva derivoiminen tuottaa kuitenkin funktioita, joiden arvot ovat rajoitettuja. Lausekkeen $f^{(n+1)}(t)$ arvolle voi siis löytää varsin yksinkertaisen ylärajan.

← Tehtävä s. 33

Toinen vihje s. 168 →

Tehtävän 19 vihje. Aluksi voisi tarkastella vaikkapa tapausta $n = 3$, jolloin laskut ovat lyhyempiä, mutta todistuksen idean kaikki elementit ovat näkyvissä.

← Tehtävä s. 33

Toinen vihje s. 169 →

Tehtävän 20 vihje. Approksimaation absoluuttinen virhe on

$$\left| f'(x) - \frac{f(x+h) - f(x)}{h} \right|.$$

Sen suuruuden arvioimiseksi voisi olla eduksi ensin korvata lauseke $f(x+h)$ sen ensimmäisen asteen Taylorin polynomilla, joka on kehitetty kohdassa x . Kun mukaan ottaa myös virhetermin, kyseessä ei ole approksimaatio vaan yhtäsuuruus.

← Tehtävä s. 34

Toinen vihje s. 169 →

Tehtävän 21 vihje. Keskusdifferenssin absoluuttinen virhe on

$$\left| f'(x) - \frac{f(x+h) - f(x-h)}{2h} \right|.$$

Kuten tehtävässä 20 lausekkeet $f(x+h)$ ja $f(x-h)$ voitaisiin korvata funktion f kohdassa x kehitetyllä Taylorin polynomilla ja virhetermeillä. Nyt kannattaisi käyttää toisen asteen Taylorin polynomia.

← Tehtävä s. 34

Toinen vihje s. 170 →

Tehtävän 22 vihje. Raja-arvon määritelmässä esiintyvän epäyhtälön $|x_n - 0| < \varepsilon$ voisi yrittää palauttaa muotoon, josta näkisi, miten suuri luvun n täytyy olla, kun luku ε on kiinnitetty.

← Tehtävä s. 45

Toinen vihje s. 170 →

Tehtävän 23 vihje. Voisit yrittää palauttaa yhtälön $|f(x) - 1| < \varepsilon$ sellaiseen muotoon, että siinä esiintyy ainakin lauseke $|x - 2|$. Voit valita luvun δ niin pieneksi,

että saat lopullisen lausekkeen pienemmäksi kuin ε . Saatat tarvita useamman eri ylärajan luvulle δ , mutta tämä ratkeaa ottamalla tarvittavista ylärajoista minimi.

← Tehtävä s. 45

Toinen vihje s. 170 →

Tehtävän 24 vihje. Voisit tehdä vastaoletuksen, että raja-arvo löytyisi, ja hyödyntää raja-arvon määritelmää ristiriidan tuottamiseksi valitsemalla sopivan pieni luku ε .

← Tehtävä s. 45

Toinen vihje s. 170 →

Tehtävän 25 vihje. Bolzanon lauseesta voisi olla tässä hyötyä.

← Tehtävä s. 45

Toinen vihje s. 170 →

Tehtävän 26 vihje. Voitaisiin tehdä vastaoletus, että raja-arvo y ei kuulu välille $[a, b]$, ja johtaa ristiriita.

← Tehtävä s. 45

Toinen vihje s. 171 →

Tehtävän 27 vihje. Voisit olettaa, että f ei ole vakiofunktio, ja etsiä ristiriidan.

← Tehtävä s. 45

Toinen vihje s. 171 →

Tehtävän 28 vihje. Voitaisiin soveltaa väliarvolauseetta.

← Tehtävä s. 45

Toinen vihje s. 171 →

Tehtävän 29 vihje. Konveksisuutta voisi lähteä selvittämään derivaatan avulla.

← Tehtävä s. 60

Toinen vihje s. 171 →

Tehtävän 30 vihje. Mallikuvan piirtäminen voisi selkiyttää tilannetta. Tässä voi soveltaa konveksin funktion määrittelyssä esiintyvää epäyhtälöä.

← Tehtävä s. 60

Toinen vihje s. 171 →

Tehtävän 31 vihje. Tulisi siis todistaa, että kun f on derivoituva ja $a < b$, pätee $f'(a) \leq f'(b)$. Tässä voisi hyödyntää pisteiden $(a, f(a))$ ja $(b, f(b))$ välistä janaa.

← Tehtävä s. 60

Toinen vihje s. 171 →

Tehtävän 32 vihje. Voitaisiin sieventää

$$\frac{a+b+c+d}{4} = \frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d,$$

jolloin Jensenin epäyhtälöä voisi soveltaa lausekkeeseen $(\frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d)^{-1}$. Sivun 53 esimerkistä voi olla apua.

← Tehtävä s. 60

Toinen vihje s. 172 →

Tehtävän 33 vihje. Väitteen epäyhtälöstä voisi ottaa logaritmifunktion puolittain, jotta tilaisuus käyttää Jensenin epäyhtälöä näkyisi selkeämmin.

← Tehtävä s. 60

Toinen vihje s. 172 →

Tehtävän 34 vihje. Tässä voisi soveltaa Jensenin epäyhtälöä sopivalle funktiolle kahdella pisteellä.

← Tehtävä s. 61

Toinen vihje s. 172 →

Tehtävän 35 vihje. Tässä voisi soveltaa Jensenin epäyhtälöä potenssifunktiolle $f(x) = x^r$. Käytettävät pisteet ja eksponentti r on mahdollista valita sopivasti.

← Tehtävä s. 61

Toinen vihje s. 172 →

Tehtävän 36 vihje. Voitaisiin valita funktiot f ja g niin, että toisesta integraalista saa tulokseksi lausekkeen $\ln(b) - \ln(a)$. Integrointiväliksi voit valita välin $[a, b]$.

← Tehtävä s. 61

Toinen vihje s. 173 →

Tehtävän 37 vihje. Kannattaa valita funktio f niin, että vasemman puolen summalauseke esiintyy kaavassa $\sum_{k=1}^n a_k f(x_k)$.

← Tehtävä s. 61

Toinen vihje s. 173 →

Tehtävän 38 vihje. Minkä suuntaisilla vektoreilla pistetulo on suurimmillaan?

← Tehtävä s. 61

Toinen vihje s. 173 →

Tehtävän 39 vihje. Yhtälön voisi palauttaa muotoon $f(x) = 0$ sopivalla funktiolla f , ja etsiä ratkaisu Newtonin menetelmällä. Eri alkuarvot voivat supeta eri juuriin.

← Tehtävä s. 78

Toinen vihje s. 173 →

Tehtävän 40 vihje. Yhtälö pitäisi muokata muotoon $x = g(x)$, jotta kiintopisteiteeraatiota voisi käyttää. Tämän voi tehdä useammalla kuin yhdellä tavalla.

← Tehtävä s. 78

Toinen vihje s. 173 →

Tehtävän 41 vihje. Suora Newtonin menetelmän soveltaminen riittää, kun valitaan sopiva funktio.

← Tehtävä s. 78

Toinen vihje s. 173 →

Tehtävän 42 vihje. Tulisi muodostaa sopiva yhtälö, johon voit soveltaa Newtonin menetelmää ja jonka ratkaisu olisi $\sqrt[3]{10}$.

← Tehtävä s. 78

Toinen vihje s. 173 →

Tehtävän 43 vihje. Newtonin menetelmä ratkaisee funktion nollakohtia, ja haluttu luku $1/a$ on esimerkiksi funktion $f(x) = x - \frac{1}{a}$ nollakohta. Valitettavasti tällä funktiolla iterointikaavaksi tulisi

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

mikä sisältää jakolaskun, vieläpä juuri sen jakolaskun, jonka laskemista yritämme kiertää.

Onneksi $f(x) = x - \frac{1}{a}$ ei ole ainoa funktio, jonka nollakohta on $x = \frac{1}{a}$.

← Tehtävä s. 78

Toinen vihje s. 173 →

Tehtävän 44 vihje. Tässä on ainakin kaksi vaihtoehtoa: Ensimmäinen ajatus voisi olla ilmaista kaava $x_2 = x_0$ Newtonin menetelmän avulla yhtälönä, jossa esiintyy vain ratkaistava piste x_0 . Tämän yhtälön voisi sitten ratkaista numeerisesti. Toinen vaihtoehto olisi hyödyntää jollakin tapaa sinifunktion parittomuutta, ja päätyä vähän yksinkertaisempaan yhtälöön.

← Tehtävä s. 79

Toinen vihje s. 174 →

Tehtävän 45 vihje. Voisi olla eduksi piirtää sopivalla ohjelmalla Newtonin menetelmän mukaisia tangentteja alkuarvosta $x = 1$ lähtien ja päätellä sen perusteella, mitä täsmällistä väitettä itse asiassa ollaan todistamassa.

← Tehtävä s. 79

Toinen vihje s. 174 →

Tehtävän 46 vihje. Kolme oleellisesti erilaista tapausta ovat $|x_0| < 1$, $|x_0| = 1$, ja $|x_0| > 1$.

← Tehtävä s. 79

Toinen vihje s. 175 →

Tehtävän 47 vihje. Voidaan edetä hyvin samalla tavalla kuin edellisessä tehtävässä.

← Tehtävä s. 80

Toinen vihje s. 175 →

Tehtävän 48 vihje. Funktion $g(x) = x^2 + c$ kiintopisteet voidaan ratkaista yhtälöstä $x^2 + c = x$.

← Tehtävä s. 80

Toinen vihje s. 175 →

Tehtävän 49 vihje. Kaavaan (19) voi sijoittaa luvut $-1, 1, 2$ pisteiksi x_j ja luvut $1, 1, 4$ kertoimien y_k paikalle. Termejä tulee yhteensä kolme kuten sivun 83 esimerkissä.

← Tehtävä s. 88

Toinen vihje s. 175 →

Tehtävän 50 vihje. Lagrangen polynomin määrittämistä varten voit sijoittaa suoraan kaavaan (19) koordinaatit $(-\pi/2, 0)$, $(0, 1)$ ja $(\pi/2, 0)$.

← Tehtävä s. 88

Toinen vihje s. 175 →

Tehtävän 51 vihje. Voisi tutkia polynomien p_{n+1} ja p_n erotusta.

← Tehtävä s. 88

Toinen vihje s. 175 →

Tehtävän 52 vihje. Tässä voisi laskea sinifunktion derivaatat, ja soveltaa kaavaa (21).

← Tehtävä s. 88

Toinen vihje s. 175 →

Tehtävän 53 vihje. Ennen koko ratkaisun etsimistä voisi miettiä, mitä käy helpommassa tapauksessa, esimerkiksi kun $N = 0$ tai $N = 1$. Näistä tapauksista voidaan myös edetä korkeamman asteen tapaukseen induktiivisesti.

← Tehtävä s. 88

Toinen vihje s. 176 →

Tehtävän 54 vihje. Voidaan sijoittaa suoraan Simpsonin säännön kaavaan (24) ja laskea.

← Tehtävä s. 103

Toinen vihje s. 176 →

Tehtävän 55 vihje. Kohdassa a) funktion arvot tulisi laskea pisteissä $x = -1$, $x = -0,5$, $x = 0$, $x = 0,5$ ja $x = 1$. Kohdat b) ja c) kannattaa laskea taulukkolaskentaohjelmalla tai koodin avulla. Kohtaan d) tarvittavat luvut voi poimia sivun 102 taulukosta. Esimerkiksi viiden desimaalin likiarvot riittävät hyvin.

← Tehtävä s. 103

Toinen vihje s. 176 →

Tehtävän 56 vihje. Kohdan a) laskussa tulisi siis laskea integraali $\int_1^2 \frac{1}{x} dx$ neljän osavälin puolisuunnikassäännöllä. Nyt $\Delta x = \frac{1}{4}$, joten saadaan arvio

$$\ln 2 \approx \frac{1}{4} \cdot \left(\frac{1}{2} f(1) + f(1,25) + f(1,5) + f(1,75) + \frac{1}{2} f(2) \right),$$

missä $f(x) = \frac{1}{x}$.

← Tehtävä s. 103

Toinen vihje s. 176 →

Tehtävän 57 vihje. Laskua voi helpottaa muistelemalla edellä esitettyä teoriaa Simpsonin säännön virhearviosta ja siinä tapahtuvista kumoutumisista.

← Tehtävä s. 104

Toinen vihje s. 177 →

Tehtävän 58 vihje. Voit käyttää suoraan puolisuunnikassäännön kaavaa, sijoittaa siihen annetut arvot ja sieventää.

← Tehtävä s. 104

Toinen vihje s. 177 →

Tehtävän 59 vihje. Luvussa esitetty Gaussin sääntö on välille $[-1, 1]$, joten joudut ensin muokkaamaan sen sopimaan välille $[2, 3]$. Tämä ei vaadi kertoimien ja pisteiden laskemista alusta asti uudelleen, vaan voi soveltaa kaavaa (27) sivulta 100.

← Tehtävä s. 104

Toinen vihje s. 177 →

Tehtävän 60 vihje. Tässä voisi edetä kuten luvussa esitettyssä esimerkissä Gaussin säännöstä kahdelle pisteelle, eli voitaisiin ratkaista oikeat arvot sopivasta yhtälöryhmästä.

← Tehtävä s. 104

Toinen vihje s. 177 →

Tehtävän 61 vihje. Tulisi siis laskea integraalille neljällä jakovälillä likiarvo I_4 , kahdella jakovälillä likiarvo I_2 , ja arvioida näiden avulla virhettä.

← Tehtävä s. 104

Toinen vihje s. 177 →

Tehtävän 62 vihje. Sievennetään väitettä hieman eteenpäin. Tulisi siis osoittaa, että approksimaatio

$$I = I_{2n} + \varepsilon_{2n} \approx I_{2n} + \frac{I_{2n} - I_n}{3} = \frac{3I_{2n}}{3} + \frac{I_{2n} - I_n}{3} = \frac{4I_{2n} - I_n}{3}$$

on Simpsonin säännön mukainen approksimaatio. Tilanne voi selkiytyä, jos laskun käy ensin läpi jollakin pienellä luvun n arvolla.

← Tehtävä s. 105

Toinen vihje s. 178 →

Tehtävän 63 vihje. Yhtälö on separoituva: se voidaan kirjoittaa muotoon $y'e^y = x$.

← Tehtävä s. 124

Toinen vihje s. 178 →

Tehtävän 64 vihje. Suuntakentän nuolten suunta riippuu siitä, onko y' positiivinen, negatiivinen vai nolla. Missä xy -tason alueissa kukin vaihtoehto toteutuu?

← Tehtävä s. 124

Toinen vihje s. 178 →

Tehtävän 65 vihje. Yhtälö on separoituva, mutta se pitää ensin kirjoittaa hieman eri muodossa.

← Tehtävä s. 124

Toinen vihje s. 178 →

Tehtävän 66 vihje. Voit ratkaista molemmat differentiaaliyhtälöt ensin erikseen, tai ratkaista ensin yhden ja sijoittaa sitten ratkaisun toiseen yhtälöön. Ensimmäinen differentiaaliyhtälö on separoituva, toinen on lineaarinen. Näille voi käyttää tekstissä esitettyjä ratkaisumenetelmiä, ja molemmista saadaan ratkaisu, jossa esiintyy vapaasti valittava vakio.

← Tehtävä s. 124

Tehtävän 67 vihje. Tässä kannattaisi ratkaista tehtävässä annettu differentiaaliyhtälö ja sijoittaa ratkaisuun tehtävässä annetut parametrit kertoimen k määrittämiseksi.

← Tehtävä s. 124

Toinen vihje s. 179 →

Tehtävän 68 vihje. Voit hyödyntää edellisessä tehtävässä selvitettyä parametrin k arvoa uudestaan, sillä se riippuu ainoastaan kappaleesta.

← Tehtävä s. 124

Toinen vihje s. 179 →

Tehtävän 69 vihje. Newtonin toisen lain mukaan $\sum \vec{F} = m\vec{a}$. Tässä liike on yhdellä suoralla, joten riittää tarkastella lukuja vektorien sijaan. Koska ilmanvastus on ainoa voima, saadaan

$$\sum F = F_d = -qv^2 = ma.$$

Tässä siis $v(t)$ on tuntematon ajan t funktio. Kiihtyvyys a puolestaan kuvaa nopeuden muutosta ajan suhteen, eli nopeuden derivaattaa ajan suhteen $v'(t)$.

Kun nopeuden lauseke on selvillä, kuljettu matka Δx saadaan laskettua (t, v) -koordinaatistosta fysikaalisena pinta-alana, eli integraalina $\Delta x = \int_{t_0}^{t_1} v(t) dt$.

← Tehtävä s. 124

Toinen vihje s. 179 →

Tehtävän 70 vihje. Lauseke $y^{\frac{2}{3}}$ on määritelty vain, kun $y \geq 0$. Alkuarvo-ongelmalla on yksi vakioratkaisu. Kun $y > 0$, yhtälö voidaan separoida ja y ratkaista integroimalla. Miten ehto $y(0) = 0$ liittyy tilanteeseen?

← Tehtävä s. 125

Toinen vihje s. 179 →

Tehtävän 71 vihje. Tässä voisi hyödyntää jollakin tapaa ratkaisujen yksikäsitteisyyttä.

← Tehtävä s. 125

Toinen vihje s. 179 →

Tehtävän 72 vihje. Voit noudattaa suoraan luvussa esiteltyä ratkaisumenetelmää toisen asteen vakiokertoimiselle yhtälölle, eli voisi aloittaa ratkaisemalla vastaavan karakteristisen yhtälön.

← Tehtävä s. 125

Toinen vihje s. 180 →

Tehtävän 73 vihje. Tulisi siis sijoittaa yhtälöön $y(x) = x^r$ ja derivoida. Tulokseksi pitäisi tulla samantyylinen karakteristinen yhtälö kuin vakiokertoimisessa tapauksessa, jolloin mahdolliset eksponentit r saa ratkaistua.

← Tehtävä s. 126

Toinen vihje s. 180 →

Tehtävän 74 vihje. Funktion y derivaatta kohdassa $x = 1$ on $y'(1) = 1 \cdot 1 = 1$. Mikä olisi siis kaavan (46) mukainen arvio koordinaateille x_1 ja y_1 ?

← Tehtävä s. 139

Toinen vihje s. 180 →

Tehtävän 75 vihje. Voidaan merkitä $h(x, y) = xy$, $\Delta x = 1$ ja $(x_0, y_0) = (1, 1)$. Lasketaan kaavoilla (47) kulmakertoimet k_1 ja k_2 :

$$k_1 = h(x_0, y_0) = h(1, 1) = 1 \cdot 1 = 1,$$

$$k_2 = h\left(x_0 + \frac{\Delta x}{2}, y_0 + k_1 \cdot \frac{\Delta x}{2}\right) = h\left(1 + \frac{1}{2}, 1 + 1 \cdot \frac{1}{2}\right) = \frac{3}{2} \cdot \frac{3}{2} = \frac{9}{4}.$$

← Tehtävä s. 139

Toinen vihje s. 180 →

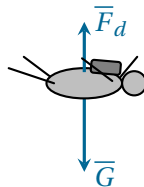
Tehtävän 76 vihje. Saatat tarvita tässä laskinta, mutta Eulerin menetelmän toteutus onnistunee suoraan sijoittamalla kaavaan (46). Nyt $h(x, y) = y$. Ensimmäisiksi arvoiksi pitäisi tulla

x	y
0,0	1
0,1	1,1
0,2	1,21
0,3	1,331

← Tehtävä s. 139

Toinen vihje s. 180 →

Tehtävän 77 vihje. Ensin olisi eduksi muodostaa toimiva fysiikkamalli. Hyppääjään vaikuttaa kaksi voimaa: painovoima $\vec{G}$ ja ilmanvastaus $\vec{F}_d$.



Positiivinen suunta voidaan sopia alaspäin, jolloin Newtonin toinen laki saa muodon

$$\begin{aligned}\Sigma \vec{F} &= m\vec{a} \\ mg - qv^2 &= ma \quad || : m \\ g - \frac{q}{m}v^2 &= a\end{aligned}$$

Nyt olisi hyvä muistaa, että kiihtyvyys on nopeuden muutosnopeus eli derivaatta ajan suhteen.

← Tehtävä s. 139

Toinen vihje s. 181 →

Tehtävän 78 vihje. Ongelma ilmenee, kun lasket pisteen (x_1, y_1) . Miksi sen y -koordinaatti on aivan liian pieni? Käänteistä menetelmää varten joudut ratkaistaan rekursioyhtälön luvun y_{n+1} suhteen.

← Tehtävä s. 140

Toinen vihje s. 181 →

Tehtävän 79 vihje. Tarvitset yhtälön todellisen ratkaisun, jonka voit laskea tulkitsemalla yhtälön epähomogeenisena lineaarisena yhtälönä. Saatat tarvita osittaisintegrointia (Liite A s. 141) oikean puolen termiä integroitaessa.

← Tehtävä s. 140

Toinen vihje s. 181 →

Tehtävän 80 vihje. Pohdinnan kannalta voisi miettiä, miten zombien ja ihmisten vuorovaikutus vaikuttaa populaatioiden kasvuun. Mieti myös, miten ihmiset reagoivat invaasioon. Eulerin menetelmää voit soveltaa samaan tyyliin kuin edellisessä jänikset ja sudet -esimerkissä.

← Tehtävä s. 140

Toinen vihje s. 181 →

Toiset vihjeet

Tehtävän 5 toinen vihje. Luku $-\frac{3}{8} = -(\frac{1}{4} + \frac{1}{8})$ on binäärilukuna $-0,011_2$. Mantissan tulisi siis olla $-1,1$ ja sopiva eksponentti on -2 , eli jos tallennettu eksponentti on e , tulisi olla $e - 1023 = -2$. Tulokseksi pitäisi tulla

[illegible]

← Tehtävä s. 20

← Vihje s. 153

Tehtävän 6 toinen vihje. Koska kulman α likiarvo on $25,3^\circ$, kulmalle α pätee $25,25^\circ \leq \alpha < 25,35^\circ$. Millä välillä kulma 4000α on? Miten tämän välin pituus suhteutuu sinifunktion jakson pituuteen?

← Tehtävä s. 20

← Vihje s. 153

Tehtävän 8 toinen vihje. Koska $\tan\left(\frac{\pi}{3}\right) = \sqrt{3}$, raja-arvo $\lim_{x \rightarrow \pi/3} L(x)$ on tangenttifunktion derivaatan arvo kohdassa $x = \frac{\pi}{3}$. Tangenttifunktion derivaatan muodoista $\frac{d}{dx}(\tan x) = 1 + \tan^2 x$ voisi olla tässä hyödyllisin.

← Tehtävä s. 21

← Vihje s. 154

Tehtävän 9 toinen vihje. Muotoa $a - \sqrt{b}$ olevan lausekkeen voi laventaa lausekkeella $a + \sqrt{b}$, jolloin saadaan

$$a - \sqrt{b} = \frac{(a + \sqrt{b})(a - \sqrt{b})}{a + \sqrt{b}} = \frac{a^2 - b}{a + \sqrt{b}}.$$

Lauseketta $1 - \sqrt{1 - \frac{v^2}{c^2}}$ voisi lähestyä tällä idealla. Approksimaatio $1 + \sqrt{1 - \frac{v^2}{c^2}} \approx 1 + 1 = 2$ ei ole ongelma, sillä se pätee hyvin monen merkitsevän numeron tarkkuudella.

← Tehtävä s. 21

← Vihje s. 154

Tehtävän 10 toinen vihje. Tulokseksi pitäisi tulla $T_5(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5$.

← Tehtävä s. 32

← Vihje s. 154

Tehtävän 11 toinen vihje. Kohdassa $x = a$ kehitetty toisen asteen Taylorin polynomi funktiolle f on muotoa

$$T_2(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2.$$

Tähän pitäisi sijoittaa $a = 4$ sekä edellä lasketut luvut $f(4)$, $f'(4)$ ja $f''(4)$. Sulklausekkeita $(x-4)$ ja $(x-4)^2$ ei kannata kertoa auki, sillä kun lopuksi lasketaan $\sqrt{5} \approx T_2(5)$, näiden arvoiksi tulee $(5-4) = 1$ ja $(5-4)^2 = 1$.

← Tehtävä s. 32

← Vihje s. 154

Tehtävän 12 toinen vihje. Taylorin polynomin muodostamista varten tarvitaan luvut $f(0)$, $f'(0)$ ja $f''(0)$. Kun $f(x) = \cos x$, luvut ovat 1, 0 ja -1 , joten lopulliseksi Taylorin polynomiksi pitäisi saada $T_2(x) = 1 - \frac{1}{2}x^2$.

← Tehtävä s. 32

← Vihje s. 154

Tehtävän 13 toinen vihje. Kun Taylorin polynomit on muodostettu, voisi olla eduksi piirtää kaikki neljä funktiota samaan koordinaatistoon. Mitä geometrisia eroja ja yhtäläisyyksiä huomaat?

Lisäksi: kummalla Taylorin polynomeista on yksinkertaisempi lauseke?

← Tehtävä s. 32

← Vihje s. 155

Tehtävän 14 toinen vihje. Taylorin polynomi on kehitetty kohdassa $a = 0$, ja likiarvoa tutkittiin kohdassa $x = 1$. Koska $\frac{d}{dx}e^x = e^x$, kaikkien kertaluokkien derivaatat ovat sama e^x . Virheen suuruus on siis

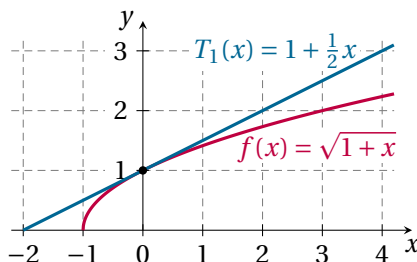
$$R_5(1) = \frac{f^{(5+1)}(t)}{(5+1)!}(1-0)^{5+1} = \frac{e^t}{6!}.$$

Tämä kaava antaa virheen tarkan arvon, mutta koska kaavassa esiintyy tuntematon luku t , olisi kätevää antaa virheelle jokin raja, jota se ei varmasti ylitä. Koska $t < 1$ ja e^x on kasvava funktio, e^t on varmasti pienempi kuin e^1 eli e . Luku $\frac{e}{6!}$ on siis yläraja virheelle. Miksi myös $\frac{3}{6!}$ on yläraja?

← Tehtävä s. 32

← Vihje s. 155

Tehtävän 15 toinen vihje. Approksimaatiot ovat kohdassa $x = 0$ kehitettyjä ensimmäisen asteen Taylorin polynomeja funktioille $f(x) = \sqrt{1+x}$ ja $g(x) = \sqrt[3]{1+x}$. Taylorin polynomien kuvaajat ovat siis funktioille f ja g pisteeseen $(0,1)$ piirrettyjä tangentteja.



Tämän perusteella voi päätellä, onko approksimaation arvo liian suuri vai liian pieni. On myös mahdollista käyttää Taylorin polynomin virhekaavaa.

← Tehtävä s. 32

← Vihje s. 155

Tehtävän 16 toinen vihje. Kun $x > 0$, $f''(x) = 0$, jolloin myös kaikki sitä korkeammat derivaatat ovat nollia. Kohdassa $x = 1$ kehitetty ”kolmannen asteen Taylorin polynomi” on siis itse asiassa ensimmäisen asteen polynomi. Itseisarvofunktion ominaisuuden kohdassa $x = 0$ selittävät, miksi tämä Taylorin polynomi toimii huonosti alueessa $x < 0$.

← Tehtävä s. 33

← Vihje s. 155

Tehtävän 17 toinen vihje. Koska funktion $f(x) = \sin x$ toinen derivaatta on $f''(x) = -\sin x$, kohdassa $x = 0$ toinen derivaatta on $f''(0) = -\sin(0) = 0$. Kohdassa $x = 0$ kehitetyn Taylorin polynomin toisen asteen termi on siis nolla, eli $T_1(x)$ ja $T_2(x)$ ovat sama polynomi. Ensimmäisen asteen polynomille $T_1(x) = x$ voidaan siis käyttää toisen asteen Taylorin polynomin virhekaavaa, jonka avulla tehtävänannon tulos on perusteltavissa.

← Tehtävä s. 33

← Vihje s. 155

Tehtävän 18 toinen vihje. Kun tarkastellaan virhetermin itseisarvoa, voidaan sieventää seuraavasti:

$$|R_n(x)| = \left| \frac{f^{(n+1)}(t)}{(n+1)!} (x-0)^{n+1} \right| = \frac{|f^{(n+1)}(t)|}{(n+1)!} |x|^{n+1}.$$

Nyt voidaan perustella, että $|f^{(n+1)}(t)| \leq 1$ ja $|x| \leq \pi$. (Miten?) Saadaan siis yläraja

$$|R_n(x)| \leq \frac{\pi^{n+1}}{(n+1)!}.$$

Riittävän pieni luvun n arvo kannattaa selvittää kokeilemalla, sillä valitettavasti epäyhtälö $\frac{\pi^{n+1}}{(n+1)!} \leq 10^{-8}$ ei ratkea analyttisesti.

← Tehtävä s. 33

← Vihje s. 156

Tehtävän 19 toinen vihje. Aiemmissa tehtävissä on laskettu, että kolmannen asteen Taylorin polynomit funktioille $f(x) = e^x$, $g(x) = \cos x$ ja $h(x) = \sin x$ ovat

$$E(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}, \quad C(x) = 1 - \frac{x^2}{2!} \quad \text{ja} \quad S(x) = x - \frac{x^3}{3!}.$$

Sievennetään nyt lauseketta $E(ix)$:

$$\begin{aligned} E(ix) &= 1 + (ix) + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} = 1 + ix + \frac{i^2 x^2}{2!} + \frac{i^3 x^3}{3!} = 1 + ix + \frac{-1 \cdot x^2}{2!} + \frac{-i \cdot x^3}{3!} \\ &= 1 - \frac{x^2}{2!} + i \left(x - \frac{x^3}{3!} \right) = C(x) + iS(x). \end{aligned}$$

Edellä käytettiin sievennystä $i^3 = i^2 \cdot i = -1 \cdot i = -i$. Todistuksen idea suuremmilla polynomin asteilla n on samanlainen: Luvun i parilliset potenssit ovat reaalityyppisiä ± 1 ja parittomat imaginaarisia lukuja $\pm i$. Kun parittomista termeistä otetaan yhteiseksi tekijäksi i , Taylorin polynomit $S(x)$ ja $C(x)$ saadaan näkyviin.

Yleisen tapauksen todistuksessa polynomin $E(ix)$ parillisista ja parittomista termeistä täytyy jotenkin pitää lukua. Parillisia voi merkitä esimerkiksi muodossa $\frac{(ix)^{2k}}{(2k)!}$ ja parittomia muodossa $\frac{(ix)^{2k+1}}{(2k+1)!}$, missä k on luonnollinen luku.

← Tehtävä s. 33

← Vihje s. 156

Tehtävän 20 toinen vihje. Taylorin polynomin kaavassa

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

esiintyvä erotus $x - a$ kuvaa approksimoitavan kohdan x poikkeamaa paikasta a , jossa polynomi on kehitetty. Nyt haluaisimme approksimoida funktion f arvoa kohdassa $x + h$ Taylorin polynomilla, joka on kehitetty kohdassa x . Poikkeaman suuruus on siis h . Voidaan kirjoittaa $f(x+h) \approx f(x) + f'(x) \cdot h$, tai virhetermin avulla tarkasti

$$f(x+h) = f(x) + f'(x) \cdot h + \frac{f''(t)}{2} h^2,$$

missä t on jokin luku lukujen x ja $x+h$ välissä. Tämän voi sijoittaa lausekkeeseen

$$\left| f'(x) - \frac{f(x+h) - f(x)}{h} \right|$$

ja sieventää.

← Tehtävä s. 34

← Vihje s. 156

Tehtävän 21 toinen vihje. Kohdassa x kehitetty toisen asteen Taylorin polynomi funktiolle f antaa virhetermeineen

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(t_1)h^3 \quad \text{ja}$$

$$f(x-h) = f(x) + f'(x)(-h) + \frac{1}{2}f''(x)(-h)^2 + \frac{1}{6}f'''(t_1)(-h)^3,$$

missä $t_2 \in [x-h, x]$ ja $t_1 \in [x, x+h]$. Nämä voisi sijoittaa keskusdifferenssin absoluuttisen virheen lausekkeeseen, jolloin sieventämisen jälkeen jäljelle pitäisi jäädä

$$\left| \frac{f'''(t_1)}{12}h^2 - \frac{f'''(t_2)}{12}h^2 \right|.$$

Lopuksi oli mahdollista hyödyntää kaikille reaaliluvuille a ja b pätevää kolmioepäyhtälöä $|a+b| \leq |a| + |b|$.

← Tehtävä s. 34

← Vihje s. 156

Tehtävän 22 toinen vihje. Soveltamalla Arkhimedeen lausetta voidaan löytää positiivinen kokonaisluku N , jolle $N > 2/\varepsilon$. Tällä valinnalla päästään haluttuun lopputulokseen.

← Tehtävä s. 45

← Vihje s. 156

Tehtävän 23 toinen vihje. Lauseketta $|f(x) - 1|$ sieventäessä päädyt lausekkeeseen, jossa esiintyy sekä lauseke $|x - 2|$ että lauseke $|x - 1|$. Tässä sinun täytyy valita δ niin, että saat epäyhtälöstä $|x - 2| < \delta$ sekä sopivan arvion lausekkeelle $|x - 2|$, mutta myös arvion, joka osoittaa, että lauseke $|x - 1|$ on sopivan kaukana nolasta. Esimerkiksi valinnalla $\delta \leq 1/2$ voit päätyä arvioon $|x - 1| > 1/2$.

← Tehtävä s. 45

← Vihje s. 156

Tehtävän 24 toinen vihje. Esimerkiksi valinnalla $\varepsilon = 1/3$ voidaan raja-arvon määritelmää hyödyntämällä pakottaa jonon häntäpää lähelle sekä lukua 1 että lukua -1 . Tästä saadaan ristiriita, sillä häntäpää ei voi samanaikaisesti olla lähellä molempia lukuja. Kaikille reaaliluvuille x ja y pätevistä kolmioepäyhtälöstä $|x + y| \leq |x| + |y|$ voi myös olla hyötyä.

← Tehtävä s. 45

← Vihje s. 157

Tehtävän 25 toinen vihje. Olisi mahdollista soveltaa Bolzanon lausetta funktiolle $g(x) = f(x) - x$, sillä jos $g(x) = 0$, niin x on funktion f kiintopiste. Muistathan, että $0 \leq f(x) \leq 1$, sillä funktion f maalijoukko on väli $[0, 1]$.

← Tehtävä s. 45

← Vihje s. 157

Tehtävän 26 toinen vihje. Jos y ei kuulu välille $[a, b]$, niin voitaisiin etsiä jokin pieni luku ε siten, että $|y - x| > \varepsilon$ kaikilla välin $[a, b]$ pisteillä x .

← Tehtävä s. 45

← Vihje s. 157

Tehtävän 27 toinen vihje. Voisi olla hyödyllistä soveltaa väliarvolauseetta tilanteeseen, jossa pisteille x ja y pätee $f(x) \neq f(y)$.

← Tehtävä s. 45

← Vihje s. 157

Tehtävän 28 toinen vihje. Voisit osoittaa väliarvolauseen nojalla, että $f(x) < f(y)$ aina, kun $x < y$.

← Tehtävä s. 45

← Vihje s. 157

Tehtävän 29 toinen vihje. Funktio on konvekksi, jos sen toinen derivaatta on ei-negatiivinen. Esimerkiksi funktion $h(x) = 1/x$ toinen derivaatta on $h''(x) = 2/x^3$, joka on positiivinen, kun $x > 0$, joten funktio on konvekksi tässä alueessa.

← Tehtävä s. 60

← Vihje s. 157

Tehtävän 30 toinen vihje. Konveksin funktion kuvaaja ei kulje kahta kuvaajan pistettä yhdistävän janan yläpuolella. Tätä voi tutkia pisteiden $(0, 0)$ ja $(3, 8)$ kautta kulkevan suoran avulla esimerkiksi valitsemalla parametri λ niin, että pisteiden $x_1 = 0$ ja $x_2 = 3$ välissä oleva piste $x = 1$ saadaan kaavasta $(1 - \lambda)x_1 + \lambda x_2$.

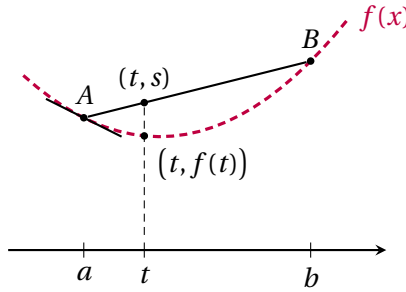
← Tehtävä s. 60

← Vihje s. 157

Tehtävän 31 toinen vihje. Väite olisi todistettu, jos saataisiin perusteltua epäyhtälö

$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b).$$

Olkoon t lukujen a ja b välissä oleva luku, ja (t, s) piste janalla AB . Mitä konveksisuuden määritelmä sanookaan lukujen s ja $f(t)$ suuruusjärjestyksestä? Miksi pätee $\frac{f(t) - f(a)}{t - a} \leq \frac{s - f(a)}{t - a}$?



← Tehtävä s. 60

← Vihje s. 157

Tehtävän 32 toinen vihje. Kun Jensenin epäyhtälöä sovelletaan lausekkeeseen $\frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d$ ja funktioon $f(x) = x^{-1}$, saadaan epäyhtälö

$$\left(\frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d\right)^{-1} \leq \frac{1}{4}(a)^{-1} + \frac{1}{4}(b)^{-1} + \frac{1}{4}(c)^{-1} + \frac{1}{4}(d)^{-1}.$$

Tätä sieventämällä on mahdollista todistaa väite.

← Tehtävä s. 60

← Vihje s. 158

Tehtävän 33 toinen vihje. Kun väitteestä otetaan puolittain logaritmi, saadaan yhtäpitävä muoto

$$\ln\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \geq \frac{1}{n} \ln(x_1 x_2 \cdots x_n).$$

Jensenin epäyhtälöä voisi soveltaa funktioon $f(x) = -\ln(x)$. Painoiksi voidaan valita sama paino $1/n$ kaikille pisteille x_k .

Tapaus, jossa jokin luvuista x_k on nolla, täytyy tarkastella eri tavalla, sillä logaritmi nolasta ei ole määritelty.

← Tehtävä s. 60

← Vihje s. 158

Tehtävän 34 toinen vihje. Voisi hyödyntää sitä, että logaritmifunktio on konkaavi, ja soveltaa Jensenin epäyhtälöä lukuihin a^p ja b^q sopivilla painokertoimilla. Tässä voi edetä samaan tapaan kuin aritmeettis-geometrisen epäyhtälön todistuksessa.

← Tehtävä s. 61

← Vihje s. 158

Tehtävän 35 toinen vihje. Hyvä valinta potenssifunktioksi olisi $f(x) = x^{\frac{p}{q}}$. Lukuja x_k ei kannata laittaa Jensenin epäyhtälöön suoraan, vaan voisi käyttää lukuja x_k^q .

← Tehtävä s. 61

← Vihje s. 158

Tehtävän 36 toinen vihje. Jensenin epäyhtälön integraaliversiota voisi soveltaa funktioihin $f(x) = x^2$ ja $g(x) = 1/x$ välillä $[a, b]$. Jensenin epäyhtälön käyttö onnistuu, koska f on konvekssi funktio.

← Tehtävä s. 61

← Vihje s. 158

Tehtävän 37 toinen vihje. Hyvä valinta on $f(x) = \sqrt{x^2 + 1}$, jonka konveksisuuden voit tarkistaa derivoimalla. Tarvitset tässä myös kaavaa $1 + 2 + \dots + n = n(n+1)/2$.

← Tehtävä s. 61

← Vihje s. 158

Tehtävän 38 toinen vihje. Tässä olisi hyödyksi pistetulon geometrinen tulkinta: Jos v ja w ovat vektoreita, niin $v \cdot w = |v||w| \cos \theta$, missä θ on vektoreiden v ja w välinen kulma.

← Tehtävä s. 61

← Vihje s. 158

Tehtävän 39 toinen vihje. Sopiva funktio asettaa Newtonin menetelmään on esim. $f(x) = e^x - 4x$, jolloin $f'(x) = e^x - 4$. Tämän jälkeen tulisi laskea laskimella tai vastaavalla työkalulla, mitä funktiota $x - f(x)/f'(x)$ iteroidessa tapahtuu. Esimerkiksi alkuarvauksen $x_0 = 0$ pitäisi toimia toisen ratkaisun löytämiseksi.

← Tehtävä s. 78

← Vihje s. 158

Tehtävän 40 toinen vihje. Yhtälön muoto $x = \frac{1}{4}e^x = g(x)$ löytää sopivalla alkuarvolla yhtälön yhden ratkaisun, mutta ei toista.

Muoto $\ln(4x) = x$ löytää toisen ratkaisun, mutta ei ensimmäistä.

← Tehtävä s. 78

← Vihje s. 159

Tehtävän 41 toinen vihje. Funktioksi voidaan valita $f(E) = E - e \sin E - M$. Alkuarvaukseksi esimerkiksi $E_0 = 1$ pitäisi toimia. Ratkaisujen lukumäärän tutkimisessa funktion kuvaajasta ja derivaatasta voi olla hyötyä.

← Tehtävä s. 78

← Vihje s. 159

Tehtävän 42 toinen vihje. Halutaan likiarvo yhtälön $x^3 = 10 \Leftrightarrow x^3 - 10 = 0$ ratkaisulle, joten Newtonin menetelmää voisi soveltaa funktion $f(x) = x^3 - 10$ nollakohdan ratkaisemiseen.

← Tehtävä s. 78

← Vihje s. 159

Tehtävän 43 toinen vihje. Funktion $f(x) = \frac{1}{x} - a$ nollakohta on sopivasti $x = \frac{1}{a}$. Tällä funktiolla iterointikaavaksi pitäisi tulla sievennyksen jälkeen $x_{n+1} = x_n(2 -$

ax_n), mikä sisältää vain kaksi kertolaskua ja yhden vähennyslaskun. Algoritmi tarvitsee luonnollisesti myös sopivan alkuarvon. Millä välillä sen tulisi olla?

← Tehtävä s. 78

← Vihje s. 159

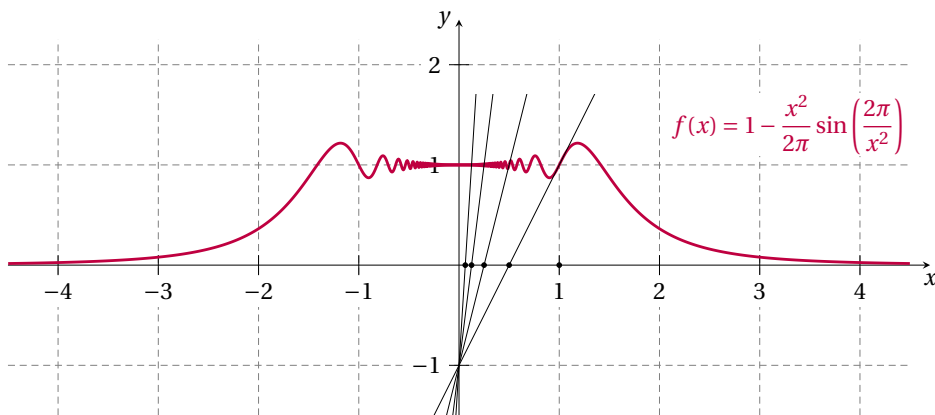
Tehtävän 44 toinen vihje. Newtonin menetelmän mukaan $x_1 = x_0 - f(x_0)/f'(x_0)$, ja vastaavasti pisteen x_2 saa esitettyä pisteen x_1 (eli myös pisteen x_0) avulla. Ensimmäinen vaihtoehto on sieventää yhtälö $x_2 = x_0$ muotoon, jossa esiintyy vain piste x_0 , ja selvittää sen ratkaisulle likiarvo Newtonin menetelmällä (sopivaan uuteen funktioon sovelletuna). Tässä tapauksessa syntyvä funktio on varsin ruma, ja sopivan alkuarvauksen löytäminen voi vaatia vähän kokeilemista.

Toinen tapa on huomata, että piste x_0 kannattaa valita niin, että $x_1 = -x_0$ (miksi?). Tämä yksinkertaistaa laskuja.

← Tehtävä s. 79

← Vihje s. 159

Tehtävän 45 toinen vihje. Alkuarvolla $x_0 = 1$ Newtonin menetelmä päättyy pisteisiin $x_1 = \frac{1}{2}$, $x_2 = \frac{1}{4}$, $x_3 = \frac{1}{8}$, ..., $x_n = 2^{-n}$, joten iteraatio suppenee lukuun $x = 0$, joka ei ole funktion nollakohta. Tämän voi perustella täsmällisesti sieventämällä Newtonin menetelmän iterointikaavaa $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$.



← Tehtävä s. 79

← Vihje s. 159

Tehtävän 46 toinen vihje. Kun $|x_0| < 1$, jono $x_0, x_0^2, (x_0^2)^2, \dots$ suppenee nopeasti kohti nollaa (joka on funktion kiintopiste). Funktion $g(x) = x^2$ derivaatta nollassa on $g'(0) = 0 < 1$, mikä sopii yhteen sen kanssa, että kyseessä on vakaa kiintopiste. Tapaukset $|x_0| > 1$ ja $x_0 = 1$ voi tarkastella vastaavasti.

← Tehtävä s. 79

← Vihje s. 160

Tehtävän 47 toinen vihje. Kiintopisteiden vakauden pitäisi tässä vaihtua toisin päin kuin edellisessä tehtävässä, ja se selviää myös derivaattaa tutkimalla.

← Tehtävä s. 80

← Vihje s. 160

Tehtävän 48 toinen vihje. Kun kiintopisteiden x_1 ja x_2 lausekkeet ovat selvillä, voisi tutkia, millä luvun c arvoilla pätee $|g'(x_1)| < 1$ tai $|g'(x_2)| < 1$.

← Tehtävä s. 80

← Vihje s. 160

Tehtävän 49 toinen vihje. Sievennyksen tarkistamista varten on hyvä huomata, että eräs tuttu toisen asteen polynomi kulkee juuri annettujen pisteiden kautta.

← Tehtävä s. 88

← Vihje s. 160

Tehtävän 50 toinen vihje. Kun lukuarvot sijoittaa Lagrangen polynomin kaavaan, vain yhdellä murtolausekkeista on nolasta poikkeava kerroin, mikä helpottaa sieventämistä. Lopputulokseksi pitäisi tulla $p_2(x) = 1 - \frac{4}{\pi^2} x^2$. Lagrangen polynomin ja Taylorin polynomin vertailussa ei tarvitse tehdä tarkkoja laskuja. Hyvän käsityksen tilanteesta saa piirtämällä polynomien ja kosinifunktion kuvaajat samaan koordinaatistoon sopivalla ohjelmalla.

← Tehtävä s. 88

← Vihje s. 160

Tehtävän 51 toinen vihje. Polynomilla $p_{n+1} - p_n$ on $n + 1$ nollakohtaa, jotka tunnetaan. Tämä antaa mahdollisuuden esittää $p_{n+1}(x) - p_n(x)$ ensimmäisen asteen tekijöiden avulla, ja tästä esityksestä voidaan sopivasti sijoittamalla ratkaista polynomi p_{n+1} .

← Tehtävä s. 88

← Vihje s. 160

Tehtävän 52 toinen vihje. Sinifunktion korkeammankin asteen derivaatat saavat arvoja vain väliltä $[-1, 1]$, joten kaavassa (21) voidaan valita $M_{n+1} = 1$. Kasvata lukua n tarpeeksi, jotta kaavan lopputulos olisi alle 10^{-5} . Tehtävässä ei kysytty pienintä lukua n , joten voit halutessasi tehdä myös epätarkkoja arvioita.

← Tehtävä s. 88

← Vihje s. 160

Tehtävän 53 toinen vihje. Voitaisiin tehdä vastaoletus, että f ei ole polynomi astetta enintään N . Tästä voisi induktion avulla näyttää, että on olemassa jono pisteitä $x_1, x_2, \dots, x_{N+2}$, joihin sovitettu Lagrangen interpolaatiopolynomi onkin astetta $N + 1$, mikä on ristiriita.

← Tehtävä s. 88

← Vihje s. 160

Tehtävän 54 toinen vihje. Virheen arvioimiseksi intergraali tulisi laskea integraalifunktion avulla. Tämä käykin suoraviivaisesti, sillä funktion e^x eräs integraalifunktio on myös e^x . Virhe on tarkan arvon ja likiarvon erotus. Virheelle voi laskea likiarvon laskimella.

← Tehtävä s. 103

← Vihje s. 160

Tehtävän 55 toinen vihje. Merkitään $f(x) = \sqrt{1-x^2}$.

- Välin pituus on 2 ja osavälejä on neljä, joten osavälin pituus on $\Delta x = 2/4 = 0,5$. Likiarvoksi saadaan

$$\pi \approx 2 \cdot \Delta x \cdot \left(\frac{1}{2}f(-1) + f(-0,5) + f(0) + f(0,5) + \frac{1}{2}f(1) \right).$$
- Nyt välejä on sata, joten $\Delta x = 2/100 = 0,02$.
- Kun pisteitä on 101, niiden väliin jää 100 väliä, joista aina kaksi vierekkäistä muodostaa Simpsonin säännön osavälin. Tällaisen osavälin pituus on siis $\Delta x = 2/50 = 0,04$. Tulisi laskea funktion f arvot kohdissa $x_1 = -1$, $x_2 = -0,98$, ..., $x_{101} = 1$. Likiarvoksi pitäisi tulla 3,13827.
- Kun pisteitä on 101, niiden väliin jää 100 väliä, joista aina kaksi vierekkäistä muodostaa Simpsonin säännön osavälin. Tällaisen osavälin pituus on siis $\Delta x = 2/50 = 0,04$.
- Sivun 102 taulukon kertoimista voidaan ottaa likiarvot luvuille $x_1, x_2, x_3, x_4, x_5, w_1, w_2, w_3, w_4, w_5$. Likiarvoksi pitäisi tulla

$$\pi \approx 2 (w_1 f(x_1) + w_2 f(x_2) + w_3 f(x_3) + w_4 f(x_4) + w_5 f(x_5)) = 3,151812 \dots$$

← Tehtävä s. 103

← Vihje s. 161

Tehtävän 56 toinen vihje. Kohdan b) virhearviossa voidaan hyödyntää sivun 97 tulosta: puolisuunnikassäännön absoluuttinen virhe on korkeintaan $\frac{M_2(b-a)}{6} (\Delta x)^2$, missä $[a, b]$ on integrointiväli, M_2 on integroitavan funktion f toisen derivaatan f'' itseisarvon suurin arvo tällä välillä, ja Δx osavälin pituus. Nyt $f(x) = \frac{1}{x}$, joten tulisi laskea toinen derivaatta $f''(x)$ ja selvittää, mikä on sen maksimiarvo välillä $[1, 2]$.

← Tehtävä s. 103

← Vihje s. 161

Tehtävän 57 toinen vihje. Simpsonin sääntö on tarkka kolmannen asteen polynomeihin asti, joten ainoastaan neljännen asteen termistä voi tulla virhettä mukaan laskuun. Riittää siis käsitellä polynomi muotoa $p(x) = cx^4$, missä c on jokin reaali-luku.

← Tehtävä s. 104

← Vihje s. 161

Tehtävän 58 toinen vihje. Kaavan sieventämiseen tarvitsee pääasiassa muistaa vain, että välipisteille pätee kaava $x_j = j/n$. Summa, jossa esiintyy luvut $f(x_j)$, sievenee tehtävänannossa annetulla summakaavalla.

← Tehtävä s. 104

← Vihje s. 161

Tehtävän 59 toinen vihje. Gaussin säännön pisteet voidaan muuttaa väliltä $[-1, 1]$ välille $[2, 3]$ käyttämällä lineaarista muunnosta $x \mapsto 0,5x + 2,5$. Muista myös laskea, mitä integraalissa tapahtuu tämän muuttujanvaihdon jälkeen, sillä myös Gaussin säännön kertoimet muuttuvat!

Tarkkaa arvoa varten tarvittaisiin funktion $x \ln(x)$ integraalifunktiota. Tämän selvittäminen lienee helpointa osittaisintegroinnin (Liite A, s. 141) avulla.

← Tehtävä s. 104

← Vihje s. 161

Tehtävän 60 toinen vihje. Riittää tutkia ainoastaan parillisen asteen monomifunktioita, jos pisteet x_k ja kertoimet w_k valitaan sopivasti, eli symmetrisesti pisteen 0 suhteen. Ratkaistavaksi yhtälöryhmäksi pitäisi tulla

$$\begin{aligned} \frac{2}{5} &= \int_{-1}^1 x^4 dx = w_1(-x_1)^4 + 0 + w_1 x_1^4 = 2w_1 x_1^4 \\ \frac{2}{3} &= \int_{-1}^1 x^2 dx = w_1(-x_1)^2 + 0 + w_1 x_1^2 = 2w_1 x_1^2 \\ 2 &= \int_{-1}^1 1 dx = w_1 + w_2 + w_1 = 2w_1 + w_2. \end{aligned}$$

← Tehtävä s. 104

← Vihje s. 161

Tehtävän 61 toinen vihje. Neljän jakovälin antamalle arviolle I_4 pätee

$$\int_0^\pi \sin x dx \approx I_4 = \frac{\pi}{4} \left(\frac{1}{2} \sin 0 + \sin \frac{\pi}{4} + \sin \frac{\pi}{2} + \sin \frac{3\pi}{4} + \frac{1}{2} \sin \pi \right).$$

Tätä voisi verrata kahden jakovälin antamaan arvioon

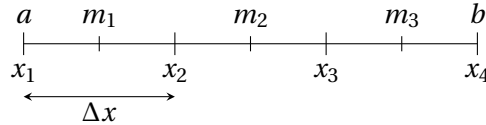
$$I_2 = \frac{\pi}{2} \left(\frac{1}{2} \sin 0 + \sin \frac{\pi}{2} + \frac{1}{2} \sin \pi \right).$$

Lukujen I_4 ja I_2 avulla voidaan laskea virheen likiarvo $\varepsilon_4 \approx \frac{I_4 - I_2}{3}$. Paljonko se poikkeaa todellisesta virheestä?

← Tehtävä s. 104

← Vihje s. 161

Tehtävän 62 toinen vihje. Tarkastellaan tapausta, jossa käytetään $n = 3$ jakoväliä, joiden leveys on Δx . Funktion arvo lasketaan siis neljässä pisteessä, olkoot ne x_1 , x_2 , x_3 ja x_4 .



Merkitään jakovälien keskipisteitä m_1 , m_2 ja m_3 . Nyt

$$I_3 = \Delta x \cdot \left(\frac{1}{2} f(x_1) + f(x_2) + f(x_3) + \frac{1}{2} f(x_4) \right)$$

ja

$$I_6 = \frac{\Delta x}{2} \cdot \left(\frac{1}{2} f(x_1) + f(m_1) + f(x_2) + f(m_2) + f(x_3) + f(m_3) + \frac{1}{2} f(x_4) \right).$$

Voisit todeta suoralla laskulla, että

$$\frac{4}{3} I_6 - \frac{1}{3} I_3 = \frac{\Delta x}{6} (f(x_1) + 4f(m_1) + 2f(x_2) + 4f(m_2) + 2f(x_3) + 4f(m_3) + f(x_4)),$$

mikä on Simpsonin säännön kaava 7 pisteelle. Samat välivaiheet toimivat myös yleisessä tapauksessa.

← Tehtävä s. 105

← Vihje s. 162

Tehtävän 63 toinen vihje. Huomaa, että $\frac{d}{dx}(e^{y(x)}) = y' \cdot e^y$. Separoitu yhtälö voidaan siis integroida puolittain muuttujan x suhteen.

← Tehtävä s. 124

← Vihje s. 162

Tehtävän 64 toinen vihje. Lauseke $y^2 x$ on nolla, kun $x = 0$ tai $y = 0$. Ratkaisukäyrien täytyy siis ylittää y -akseli vaakasuoraan. Derivaatan merkki riippuu x -koordinaatista.

← Tehtävä s. 124

← Vihje s. 162

Tehtävän 65 toinen vihje. Yhtälön oikean puolen voisi jakaa tekijöihin muodossa $(y + 1)(x + 1)$. Tämän jälkeen voisi etsiä yhden vakioratkaisun ja ratkaista loput separoimalla.

← Tehtävä s. 124

← Vihje s. 162

Tehtävän 67 toinen vihje. Tehtävän yhtälöä voi ajatella joko separoituvana tai lineaarisena epähomogeenisena yhtälönä. Kumpikin ratkaisumenetelmä toimii. Kun olet selvittänyt ratkaisufunktion $T(t)$ kaavan, esitä tehtävässä annetut oletukset lämpötilan suuruudesta kahtena yhtälönä. Näistä voit määrittää ratkaisussa esiintyvän vakioparametrin ja kertoimen k .

← Tehtävä s. 124

← Vihje s. 162

Tehtävän 68 toinen vihje. Vakio C ei ole sama kuin edellisessä tehtävässä, vaan se pitää selvittää alkuehdosta $T(0) = M + C \cdot 1$. Ympäristön lämpötila M voidaan ratkaista yhtälöistä $T(0) = 200$ ja $T(540) = 100$.

← Tehtävä s. 124

← Vihje s. 162

Tehtävän 69 toinen vihje. Edellisen vihjeen perusteella ratkaistavana on differentiaaliyhtälö

$$-q(v(t))^2 = mv'(t).$$

Vakioratkaisu $v(t) = 0$ ei sovi alkuarvoon $v(0) = v_0 > 0$. Yhtälö on separoituva, ja muodon

$$-\frac{v'(t)}{v(t)^2} = \frac{q}{m}.$$

voi integroida puolittain ajan suhteen.

← Tehtävä s. 124

← Vihje s. 163

Tehtävän 70 toinen vihje. Vakioratkaisu on $y(x) \equiv 0$, ja jos $y(x) \neq 0$, separoimalla saadaan $y(x) = (x + C)^3$. Valinnalla $C = 0$ voitaisiin rakentaa paloittain määritelty ratkaisu

$$y(x) = \begin{cases} x^3, & \text{kun } x \geq 0 \\ 0, & \text{kun } x < 0. \end{cases}$$

Miksi se toteuttaa alkuarvo-ongelman? Ratkaisuja on vielä muitakin. Millainen ratkaisu voitaisiin rakentaa, jos lausekkeessa $(x + C)^3$ olisi $C < 0$?

← Tehtävä s. 125

← Vihje s. 163

Tehtävän 71 toinen vihje. Millaisia vakioratkaisuja yhtälöllä on? Yksikäsitteisyyslauseen perusteella funktion y kuvaaja ei voi leikata vakioratkaisujen kuvaajaa. Tästä voisi päätellä, että $y(x)$ kuuluu halutulle välille.

← Tehtävä s. 125

← Vihje s. 163

Tehtävän 72 toinen vihje. Karakteristisesta yhtälöstä pitäisi tulla diskriminantiksi nolla, joten ollaan ratkaisutapauksessa kolme. Ratkaisu on muotoa $y(x) = C_1 e^{2x} + C_2 x e^{2x}$, missä vakiot C_1 ja C_2 määräytyvät alkuehdoista.

← Tehtävä s. 125

← Vihje s. 163

Tehtävän 73 toinen vihje. Voit rajoittua tapaukseen, jossa johtamallasi yhtälöllä eksponentille r on kaksi reaalista ratkaisua. Niiden avulla voidaan tuottaa alkuperäiselle yhtälölle kaksi lineaarisesti riippumatonta ratkaisufunktiota, joiden lineaarikombinaatio muodostaa yleisen ratkaisun.

Tapauksessa $a = 3$ ja $b = -3$ pitäisi käydä juuri näin, sillä eksponentille r tulee toisen asteen yhtälö $r^2 + 2r - 3 = 0$, jolle löydät kaksi reaalista ratkaisua.

← Tehtävä s. 126

← Vihje s. 163

Tehtävän 74 toinen vihje. Tuloksiksi pitäisi tulla $y(2) \approx 2$ ja $y(3) \approx 6$. Nämä ovat huomattavan kaukana oikeista arvoista.

← Tehtävä s. 139

← Vihje s. 163

Tehtävän 75 toinen vihje. Ensimmäisen vihjeen tavoin jatkaen kulmakertoimiksi pitäisi tulla $k_3 = \frac{51}{16}$ ja $\frac{134}{16}$, ja lopulliseksi arvioksi $y(2) \approx \frac{35}{8} = 4,375$.

← Tehtävä s. 139

← Vihje s. 164

Tehtävän 76 toinen vihje. Kannattaa ehkä ensin ratkaista yleinen tapaus, sillä menetelmän antama arvio tulee suppenemaan varsin näppärään (ja ehkä tuttuun) muotoon. Arvot $y_1, y_2, y_3, \dots, y_N$ muodostavat geometrisen jonon.

← Tehtävä s. 139

← Vihje s. 164

Tehtävän 77 toinen vihje. Kiihtyvyys $a(t)$ on nopeuden $v(t)$ derivaatta ajan suhteen, joten edellisen vihjeen yhtälö saa muodon

$$v'(t) = g - \frac{q}{m} [v(t)]^2.$$

Tässä g , k ja m ovat vakioita, joten kysessä on differentiaaliyhtälö funktiolle v alkuarvolla $v(0) = 0 \left(\frac{\text{m}}{\text{s}}\right)$. Sen voi ratkaista Eulerin menetelmällä esimerkiksi tau-lukkolaskentaohjelmalla, jolloin yhteen soluun tulisi aina nopeus tietyinä ajanhetkenä.

Ensimmäisten sekuntien tuloksiksi pitäisi tulla

$$\begin{aligned} v(0) &= 0 \\ v(1) &= 9,81 \\ v(2) &= 19,286\dots \\ v(3) &= 27,806\dots \\ &\vdots \end{aligned}$$

← Tehtävä s. 139

← Vihje s. 164

Tehtävän 78 toinen vihje. Käänteisistä menetelmästä pitäisi tulla rekursiokaava $y_{n+1} = y_n/101$. Voit verrata lopullisia arvoja pisteessä $x = 1$ esimerkiksi ottamalla logaritmifunktion molemmista.

← Tehtävä s. 140

← Vihje s. 165

Tehtävän 79 toinen vihje. Yhtälö kannattaa kertoa puolittain lausekkeella e^{-x} , jotta pääset tekemään jotakin lausekkeelle $y' - y$. Oikealle puolelle jää ikävä lauseke $e^{-x} \sin(x)$, mutta tämän integraalifunktion voi määrittää osittaisintegroimalla kahdesti peräkkäin ja ratkaisemalla integraalin muodostuneesta yhtälöstä.

← Tehtävä s. 140

← Vihje s. 165

Tehtävän 80 toinen vihje. Ihmisten ja zombien kohdatessa osa ihmisistä muuttuu zombeiksi, ja lisäksi ihmiset vähentävät zombien lukumäärää taistelemalla. Tekijä $1 + \gamma Z(t)$ voisi ilmaista ihmisten taisteluhalukkuutta. Eulerin menetelmässä rekursiokaavat ovat

$$H_{n+1} = H_n - \beta H_n Z_n \Delta t \quad \text{ja} \quad Z_{n+1} = Z_n + (\beta - \alpha (1 + \gamma Z_n)) H_n Z_n \Delta t.$$

← Tehtävä s. 140

← Vihje s. 165

Lähteet

Kirjan teksti on ladottu $\text{\LaTeX}$ -järjestelmällä ja kuvat piirretty pääsääntöisesti sen *TikZ*-paketilla. Joidenkin kuvien piirtämisessä on hyödynnetty myös ohjelmistoa GeoGebra 6 [2] ja Python-ohjelmointikielen Matplotlib-kirjastoa.

Numeriikkaa käsittelevät luvun 1 pääasiallisena inspiraationa ja lähteenä ovat pitkät keskustelut matematiikan opettajan, sotatieteen tohtori Esa Lapin kanssa. Termistön osalta on pyritty seuraamaan lukion pitkän matematiikan vanhan Pyramidi-oppikirjasarjan osaa Numeeriset menetelmät [8].

Lukujärjestelmien historiasta löytyy runsaasti lisätietoa Graham Fleggin perusteellisesta teoksesta *Lukujen historia* [1].

Liukulukustandardi binary64 sisältyy IEEE 754 -standardiin [6], jota käsittelevä Wikipedia-artikkeli [15] on varsin kattava.

Laajempaa käsittelyä differentiaalilaskennan täsmällisistä määritelmistä löytyy klassikkoteoksista [11] ja [12], tai uudemmassa oppikirjasta [7].

Taylorin polynomin virhekaavan todistuksen (Liite D, s. 145) kattavampi versio löytyy teoksesta [11] osioista 9.2 sekä teoksesta [12] osioista 15.2.

Differentiaaliyhtälöitä käsittelevän luvun 8 päälähteenä on ollut David Sánchezin teos [13].

Luvuissa 7 ja 8 käsiteltyjen numeeristen menetelmien tarkempaan teoriapohjaan voi tutustua esimerkiksi teoksessa [9].

Kuvalähteet

Luvun 1 aloittava simulaatiokuvan lähde on NASA [25], ja sivun 10 kuva Eiffel tornista on Wikimedia Commons -sivulta [24].

Kuva Brook Taylorista on Wikimedia Commons -sivulta [22], Isaac Newtonin kuva sivustolta [23], Joseph-Louis Lagrangen kuva sivustolta [19], Thomas Simpsonin kuva sivustolta [17], Carl Friedrich Gaussin kuva on sivustolta [20], Leonhard Eulerin kuva sivustolta [21] ja kuva Ernst Lindelöfistä museoviraston kokoelmasta [18].

Tehtävien lähteet

Tehtävä 6. Tämän viehättävän tehtävän idean tekijät kuulivat Esa Lapilta.

Tehtävä 8. Tehtävä on syksyn 2005 pitkän matematiikan ylioppilaskokeen tehtävä 13. Alkuperäisessä tehtävässä pyydettiin tutkimaan arvoja $n = 1, 2, 3, 4, 5$.

Tehtävä 20. Tehtävä on esimerkkinä oppikirjan [8] osiossa 6.1.

Tehtävä 21. Tämäkin tehtävä on esimerkkinä oppikirjan [8] osiossa 6.1.

Tehtävä 45. Tehtävänannon funktio on muokattu esimerkistä, jonka Juha Arpiainen kehitti ollessaan opiskelijana Päivölän opiston matematiikkalinjalla.

Tehtävä 61. Olemme kuulleet tämän ajatuksen Esa Lapilta.

Tehtävä 62. Tämänkin tehtävän lähteenä ovat keskustelut Esa Lapin kanssa.

Tehtävä 70. Tehtävä on esimerkki kirjan [13] luvusta 2.

Viitteet

- [1] FLEGG, G.: *Lukujen historia – Sormilla laskemisesta tietokoneisiin*, Art House, 2002.
- [2] INTERNATIONAL GEOGEBRA INSTITUTE: *GeoGebra Classic 6*, 2016.
- [3] CHADWICK, S.G. ja S.J. HAAKE: *The Drag Coefficient of Tennis Balls*, Engeneering of Sport Conference Sydney, June 2000.
- [4] HAFELE, J.C. ja R.E. KEATING: *Around-the-World Atomic Clocks: Observed relativistic Time Gains*, Science, Vol 177, 4044, 1972.
- [5] HUNTER, J. D.: *Matplotlib: A 2D Graphics Environment*, Computing in Science & Engineering, vol. 9, no. 3, 90–95, 2007.
- [6] IEEE STANDARDS ASSOCIATION: *IEEE Standard for Floating-Point Arithmetic*, IEEE Computer Society, 2019.
- [7] HARJULEHTO, P., R. KLÉN ja M. KOSKENOJA *Analyysiä reaaliluvuilla*, Gaudeamus, 2023.
- [8] HEMMO, K., E. LAPPI ja R. LUNDAHL: *Pyramidi, numeeriset menetelmät*, Kustannusosakeyhtiö Tammi, 2002.
- [9] KRESS, R.: *Numerical Analysis*, Graduate Texts in Mathematics, Springer-Verlag, 1998.
- [10] THE MPMATH DEVELOPMENT TEAM: *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)*, 2023.
<http://mpmath.org/>
- [11] MYRBERG, L.: *Differentiaali- ja integraalilaskenta korkeakouluja varten, osa 1*, Kirjayhtymä Oy, 1996.
- [12] MYRBERG, L.: *Differentiaali- ja integraalilaskenta korkeakouluja varten, osa 2*, Kustannusosakeyhtiö Tammi, 2001.
- [13] SÁNCHEZ, D.: *Ordinary Differential Equations, A Brief Eclectic Tour*, The Mathematical Association of America, 2002.

- [14] TILVIS, V. ja T. TÖRMÄ: *Pitkän matematiikan lisäsivut 5*, Viipurin Reaalikoulu Oy, 2025.
- [15] WIKIPEDIA: *IEEE 754*. Luettu 9.2.2025.
https://en.wikipedia.org/wiki/IEEE_754
- [16] WOLFRAM RESEARCH, INC.: *Wolfram|Alpha*
<https://www.wolframalpha.com/>

Kuvaviitteet

- [17] MACTUTOR: *Thomas Simpson*. Luettu 22.3.2025. Public Domain.
<https://mathshistory.st-andrews.ac.uk/Biographies/Simpson/>
- [18] NYBLIN, D.: Valokuva *Matematiikko Ernst Lindelöf vasta valmistuneena filosofian tohtorina*. Museovirasto, Historian kuvakokoelma. CC-BY-4.0-lisenssi. Luettu 9.4.2025.
<https://finna.fi/Record/museovirasto.8F98787F9F214903052B022904B1F1EE>
- [19] STORE NORSKE LEKSIKON: *Joseph Louis Lagrange*. Public Domain. Luettu 31.3.2025.
https://snl.no/Joseph_Louis_Lagrange
- [20] WIKIMEDIA COMMONS: Kuva *Carl Friedrich Gauss*. Public Domain. Luettu 22.3.2025.
https://commons.wikimedia.org/wiki/File:Carl_Friedrich_Gauss.jpg
- [21] WIKIMEDIA COMMONS: Kuva *Jakob Emanuel Handmann: Portrait of Leonhard Euler (1707-1783)*. Public Domain. Luettu 1.4.2025.
[https://commons.wikimedia.org/wiki/File:Leonhard_Euler_-_Jakob_Emanuel_Handmann_\(Kunstmuseum_Basel\).jpg](https://commons.wikimedia.org/wiki/File:Leonhard_Euler_-_Jakob_Emanuel_Handmann_(Kunstmuseum_Basel).jpg)
- [22] WIKIMEDIA COMMONS: Kuva *Taylor Brook Goupy NPG*. Public Domain. Luettu 15.9.2024.
https://commons.wikimedia.org/wiki/File:Taylor_Brook_Goupy_NPG.jpg
- [23] WIKIMEDIA COMMONS: Kuva *Portrait of Isaac Newton (1642-1727)*. Public Domain. Luettu 29.3.2025.
[https://en.wikipedia.org/wiki/Isaac_Newton#/media/File:Portrait_of_Sir_Isaac_Newton,_1689_\(brightened\).jpg](https://en.wikipedia.org/wiki/Isaac_Newton#/media/File:Portrait_of_Sir_Isaac_Newton,_1689_(brightened).jpg)
- [24] WIKIMEDIA COMMONS: Kuva *Tour Eiffel*. Public Domain. Luettu 26.1.2025.
[https://commons.wikimedia.org/wiki/File:Tour_Eiffel_Wikimedia_Commons_\(cropped\).jpg](https://commons.wikimedia.org/wiki/File:Tour_Eiffel_Wikimedia_Commons_(cropped).jpg)
- [25] WIKIMEDIA COMMONS: Kuva *Tracking a Superstorm(8970258657)*. NASA sallii oppikirjakäytön. Luettu 26.1.2025.
https://commons.wikimedia.org/wiki/File:Tracking_a_Superstorm_%288970258657%29.png