

Luku 6

Differentiaaliyhtälöt

*Nous avons calculé plus de 120 trajectoires différentes,
travail immense qui a exigé plus de 4500 heures.*

*Quand on est suffisamment exercé,
on calcule environ trois points (R, z) par heure.¹*

C. Störmer, 1907

In general, books on numerical ODEs²

fall into two categories: pre-Dahlquist and post-Dahlquist.

The first set is nowadays of mainly historical and antiquarian significance.

A. Iserles [17]

6.1 Taustaa

Difyhtälöitten numeerisia menetelmiä tutkittiin jonkin verran jo 1700-luvulla, ja yksinkertaisin alkuarvotehtävien numeerinen menetelmä on luonnollisesti nimeltään Eulerin menetelmä. Kiinnostus niihin oli kuitenkin varsin vähäistä, koska käsin laskien menetelmät ovat erittäin työläitä, kuten Störmerin kommentista selviää.

Tietokoneitten keksimisen myötä tutkimus kuitenkin nopeasti laajeni, ja yllämainittu ruotsalainen Germund Dahlquist oli yksi niistä jotka 50-luvun lopulla aloittivat tämän difyhtälöitten numeeristen menetelmien tutkimuksen modernin kauden. Seuraavat kirjat ovat hyviä ja perusteellisia johdatuksia alkuarvotehtävien numeerisiin menetelmiin [5, 8, 10, 11, 17]. Hyviä johdantoja difyhtälöitten perusteoriaan ovat muun muassa [2, 14, 15, 16].

¹ Olen laskenut yli 120 eri ratkaisua, valtava työ joka on vienyt yli 4500 tuntia.
Riittävän harjaantunut laskija laskee noin kolme pistettä tunnissa.

Störmer ratkaisi kahden toisen kertaluvun difyhtälön systeemiä, jotka mallittivat varattujen hiukkasten ratoja niitten saapuessa avaruudesta maan magneettikenttään.

²ODE = ordinary differential equation.

Palautetaan aluksi mieleen muutama perusasia difyhtälöistä. Tarkastellaan alkuarvototehtävää

$$\begin{cases} x'(t) = f(t, x(t)) \\ x(t_0) = x_0 \end{cases} \quad (6.1)$$

Olkoon $\Omega \subset \mathbb{R}^2$ jokin sopiva osajoukko ja $f : \Omega \rightarrow \mathbb{R}$. Oletetaan lisäksi että f , $\partial f / \partial t$ ja $\partial f / \partial x$ ovat jatkuvia Ω :ssa; merkitään $f \in C^1(\Omega)$.

Lause 6.1. *Oletetaan, että*

(i) Ω on avoin ja $(t_0, x_0) \in \Omega$

(ii) $f \in C^1(\Omega)$

Tällöin on olemassa positiiviset a, b ja kuvaus $z : [t_0 - a, t_0 + a] \rightarrow \mathbb{R}$ siten, että

1. $\Omega_{a,b} = [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b] \subset \Omega$

2. $(t, z(t)) \in \Omega_{a,b}$ kaikilla $t \in [t_0 - a, t_0 + a]$

3. z on tehtävän (6.1) yksikäsitteinen ratkaisu.

Tunnetusti yleisesti ottaen pitää tyytyä paikalliseen ratkaisuun: kun f on epälinaarinen x :n suhteen niin on helppo keksiä esimerkkejä joissa ratkaisu ei ole määritelty kaikilla t :n arvoilla. Jatkossa ei kuitenkaan huolehdi tästä vaan oletetaan nyt aina että tarkasteltavalla difyhtälöllä on yksikäsitteinen ratkaisu sillä välillä jota halutaan tarkastella.

Lauseen 6.1 todistus perustuu *Picardin iteraatioon* joka voidaan tulkita kiintopisteiteraatioksi. Siis difyhtälöitten olemassaolo- ja yksikäsitteisyyslause on eräs sovellus Banachin kiintopistelauseesta. Kuten muistetaan, niin kiintopisteiteraatio on konstruktiiivinen: siinähan muodostetaan jono joka suppenee haluttuun ratkaisuun. Periaatteessa siis Picardin iteraatiota voitaisiin käyttää myös numeerisena menetelmänä. Numeerisessa laskennassa pelkkä suppeneminen ei kuitenkaan riitä, suppenemisen pitäisi lisäksi olla riittävän nopeaa. Onkin osoittautunut että käytännöllisiä numeerisia menetelmiä pitää etsiä toisaalta.

6.2 Kaksi yksinkertaista menetelmää

Tehtävää (6.1) voidaan lähestyä monella tavalla, ja katsotaan seuraavassa 2 tapaa. Otetaan kuitenkin ensin käyttöön muutama merkintä. Olkoon h jokin positiivinen parametri jota sanotaan *askelpituudeksi* ja olkoon $t_n = t_0 + nh$. Olkoon x tehtävän (6.1) ratkaisu ja haluttaisiin pisteissä t_n laskea x :lle approksimaatiot x_n ; siis $x_n \approx x(t_n)$. Huomaa, että koska alkuarvo on annettu, niin $x_0 = x(t_0)$. Tästä saadaan

Ongelma 6.1. Miten laskea x_n siten, että virheet $x(t_n) - x_n$ olisivat mahdollisimman (tai riittävän) pienet? *

Askelpituus h määrittää kuinka tiheään ratkaisua x halutaan approksimoida. Tavoitteena on siis että kun $h \rightarrow 0$ niin myös virheet lähestyisivät nollaa.

Ensimmäinen idea on todeta, että numeerinen integrointi on erikoistapaus tehtävästä 6.1. Jos f ei riipu x :stä, niin selvästi

$$x(t) = x(t_0) + \int_{t_0}^t f(s)ds = x_0 + \int_{t_0}^t f(s)ds$$

Approksimoidaan integraalia trapetsimenetelmällä:

$$x_n = x_0 + \frac{h}{2} (f(t_0) + 2f(t_1) + \cdots + 2f(t_{n-1}) + f(t_n))$$

Mutta nyt huomataan, että on tarpeetonta aloittaa integrointi yhä uudestaan ajan hetkestä t_0 vaan voidaan kirjoittaa

$$\begin{aligned} x_{n+1} &= x_0 + \frac{h}{2} (f(t_0) + 2f(t_1) + \cdots + 2f(t_n) + f(t_{n+1})) \\ &= x_0 + \frac{h}{2} (f(t_0) + 2f(t_1) + \cdots + 2f(t_{n-1}) + f(t_n)) + \frac{h}{2} (f(t_n) + f(t_{n+1})) \\ &= x_n + \frac{h}{2} (f(t_n) + f(t_{n+1})) \end{aligned}$$

Mutta nythän voisi arvella, että tämä olisi järkevä myös siinä tapauksessa, että f riippuu x :stä:

$$x_{n+1} = x_n + \frac{h}{2} (f(t_n, x_n) + f(t_{n+1}, x_{n+1}))$$

Tätä(kin) on tapana sanoa *trapetsimenetelmäksi*. Huonona puolena tässä on se, että jos f on epälineaarinen x :n suhteen, niin joka askeleella x_{n+1} pitää ratkaista epälineaarisesta yhtälöstä.

Toinen tapa edetä on lähteä liikkeelle geometrisesti. Olkoon z jokin tehtävän 6.1 ratkaisu ja olkoon $(t_*, z(t_*))$ jokin piste ratkaisukäyrällä. Tällöin z :n määrittämän käyrän tangentti pisteessä $(t_*, z(t_*))$ on

$$x - z(t_*) = z'(t_*)(t - t_*) = f(t_*, z(t_*))(t - t_*)$$

Mutta nyt jos annetaan mikä tahansa tason piste $p = (\bar{t}, \bar{x})$, niin nyt tiedetään, että tämän pisteen kautta kulkevan ratkaisun tangentti on

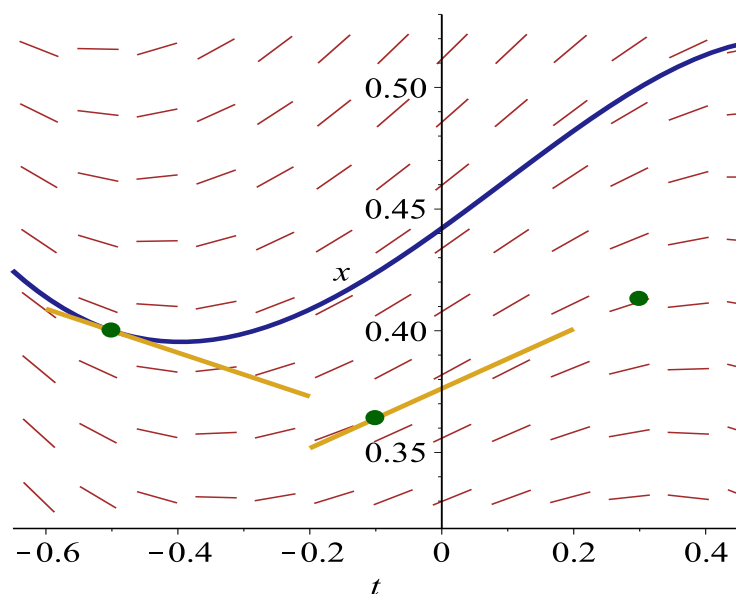
$$x - \bar{x} = f(p)(t - \bar{t}) \tag{6.2}$$

vaikka itse ratkaisua ei tunnetakaan.

Määritelmä 6.1. Yhtälön (6.2) määrittämä suoraparvi on *difyhtälön* (6.1) suuntakenttä

Koska tangentti approksimoi itse käyrää hyvin sivuamispisteen lähistöllä niin tämä johtaa yksinkertaiseen menetelmään. Lähdetään liikkeelle alkupisteestä (t_0, x_0) . Valitaan aika-askel h ja lasketaan seuraava piste etenemällä tangenttia pitkin:

$$t_1 = t_0 + h \quad \text{ja} \quad x_1 = x_0 + f(t_0, x_0)(t_1 - t_0) = x_0 + hf(t_0, x_0)$$



Kuva 6.1: Suuntakenttä ja kaksi askelta Eulerin menetelmällä.

Tämän jälkeen lasketaan tangentti pisteessä (t_1, x_1) ja edetään tämän suuntaisesti. Tätä on tapana kutsua *Eulerin menetelmäksi*. Kuvassa 6.1 on otettu 2 askelta kun $f = x^2 - t^2$, $h = 0.4$ ja $(t_0, x_0) = (-0.5, 0.4)$.

Algoritmi 8 EULERIN MENETELMÄ

Input: funktio f , alkuhetki t_a , loppuhetki t_b , alkuarvo x_a , askelten määrä N

Output: pisteet $p^n = (t_n, x_n)$, $0 \leq n \leq N$

$h := (t_b - t_a)/N$; $t_0 := t_a$; $x_0 := x_a$;

for $n = 0$ to $N - 1$ **do**

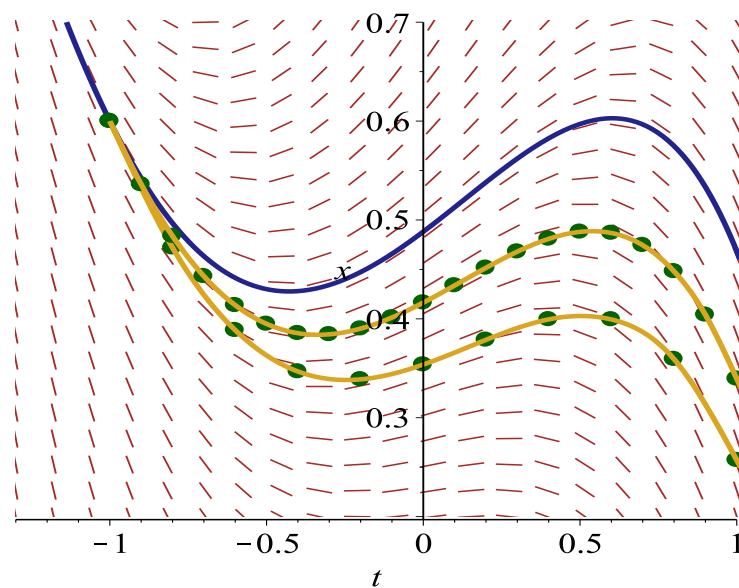
$t_{n+1} := t_n + h$;

$x_{n+1} := x_n + h f(t_n, x_n)$

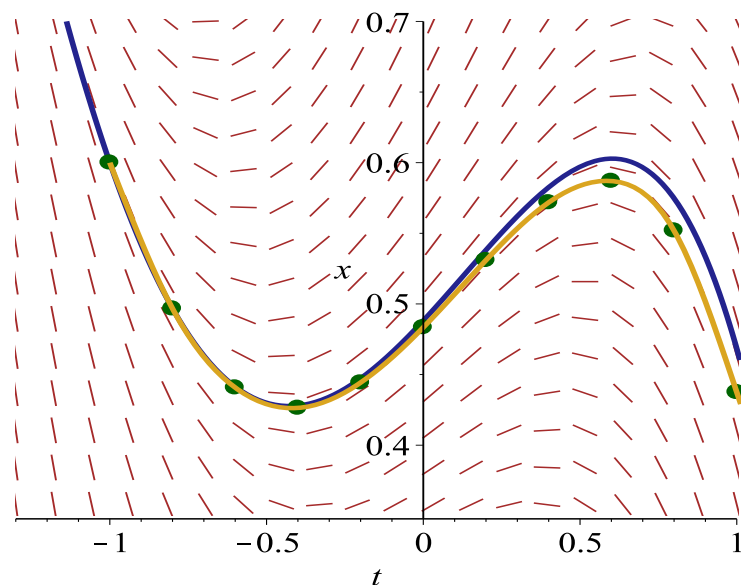
end for

Esimerkki 6.1. Olkoon $x' = x^2 - t^2$ ja valitaan alkuarvoksi $x(-1) = 0.6$. Kuvassa 6.2 on 2 Eulerin menetelmällä saatua numeerista ratkaisua 2 eri aika-askeleella. Pienemmällä aika-askeleella saadaan odotetusti tarkempi ratkaisu, mutta virhe on joka tapauksessa iso. Numeerisen ratkaisun antamat pisteet on luonnollisesti interpoloitu kuutio splineä käyttäen. Kuvassa 6.3 puolestaan on saman tehtävän ratkaisu trapetsimenetelmällä. Selvästi nyt saadaan paljon parempi approksimaatio. ★

Mutta suppeneeko Eulerin menetelmä todella? Määritellään ensin tarkemmin mitä ylipäätään tarkoitetaan menetelmän suppenemisella. Tarkastellaan yleistä yhtälöä (6.1) ja halutaan numeerinen ratkaisu välille $[t_0, t_0 + T]$. Valitaan jokin N ja asetetaan $h = T/N$. Asetetaan edelleen alkuarvoksi $x_0 = x(t_0)$ ja lasketaan arvot x_n , $1 \leq n \leq N$ jollain menetelmällä.



Kuva 6.2: Esimerkin 6.1 ratkaisu ja 2 numeerista ratkaisua käyttäen askelpituuksia $h = 0.2$ ja $h = 0.1$.



Kuva 6.3: Esimerkin 6.1 ratkaisu ja numeerinen ratkaisu käyttäen trapetsimenetelmää ja askelpituutta $h = 0.2$. Selvästi näinkin isolla askeleella saadaan huomattavasti tarkempi tulos kuin Eulerin menetelmällä.

Määritelmä 6.2. *Numeerisen ratkaisun virhe on $\mathcal{E}_h = \max_{0 \leq n \leq N} |x(t_n) - x_n|$. Menetelmä suppenee, jos $\mathcal{E}_h \rightarrow 0$, kun $h \rightarrow 0$.*

Esimerkki 6.2. Tarkastellaan tehtävää $x' = -x$, $x(0) = 1$ jonka ratkaisu on $x(t) = e^{-t}$ ja sovelletaan tähän Eulerin menetelmää. Tällöin

$$x_{n+1} = x_n + hf(t_n, x_n) = (1 - h)x_n \quad \Rightarrow \quad x_n = (1 - h)^n x_0 = (1 - h)^n$$

Olkoon $\varepsilon_n = x(t_n) - x_n$ jolloin

$$\begin{aligned} \varepsilon_n &= e^{-nh} - (1 - h)^n = (e^{-h})^n - (1 - h)^n \\ &= (e^{-h} - 1 + h)(e^{-(n-1)h} + e^{-(n-2)h}(1 - h) + \dots + e^{-h}(1 - h)^{n-2} + (1 - h)^{n-1}) \end{aligned}$$

Koska $e^{-h} \leq 1$ ja $|1 - h| \leq 1$ kun h on riittävän pieni niin saadaan

$$|\varepsilon_n| \leq n(e^{-h} - 1 + h)$$

Toisaalta Taylorin lauseen perusteella on olemassa $0 \leq \beta \leq h$ siten että

$$e^{-h} - 1 + h = \frac{1}{2} e^{-\beta} h^2 \quad \Rightarrow \quad |\varepsilon_n| \leq \frac{n}{2} e^{-\beta} h^2 \leq \frac{n}{2} h^2$$

Jos sitten valitaan $h = T/N$, niin

$$\mathcal{E}_h = \max_{0 \leq n \leq N} |\varepsilon_n| \leq \frac{N}{2} h^2 = \frac{T}{2} h = \mathcal{O}(h)$$

Siis Eulerin menetelmä suppenee, ainakin tämän esimerkin tapauksessa, mutta suppeneminen on hyvin hidasta. ★

6.3 Erilaisia menetelmäluokkia

Sekä trapetsimenetelmä että Eulerin menetelmä ovat varsin epätarkkoja joten haluttaisiin parempia menetelmiä. Tähän on kaksi erilaista tapaa jotka molemmat ovat hyödyllisiä. Olkoon taas $t_n = t_0 + nh$ ja $x_n \approx x(t_n)$. Edellä tarkastellut menetelmät käyttivät vain viimeisintä laskettua arvoa: x_{n+1} riippui x_n :sta, mutta ei aiemmin lasketuista arvoista. *Moniaskelmenetelmissä* käytetään myös näitä varhaisempia arvoja ja pyritään siten parantamaan tarkkuutta. Yleinen ℓ -askelinen menetelmä voidaan kirjoittaa muodossa

$$\sum_{j=0}^{\ell} a_j x_{n+j} = h \sum_{j=0}^{\ell} b_j f(t_{n+j}, x_{n+j}) \quad (6.3)$$

Parametrit a_j ja b_j siis määrittävät menetelmän. Luonnollisesti menetelmä ei muutu jos kaikkia kertoimia kerrotaan samalla luvulla, ja on tapana normalisoida asettamalla joko $a_\ell = 1$ tai $\sum_j b_j = 1$. Näillä merkinnöillä

(E) Eulerin menetelmällä $\ell = 1$, $a_1 = -a_0 = b_0 = 1$ ja $b_1 = 0$

(T) trapetsimenetelmällä $\ell = 1$, $a_1 = -a_0 = 1$ ja $b_1 = b_0 = 1/2$

Sanotaan, että menetelmä on *eksplisiittinen*, jos $b_\ell = 0$ ja *implisiittinen*, jos $b_\ell \neq 0$. Implisiittiset menetelmät ovat siitä hankalia, että jos f on epälineaarinen x :n suhteen, niin joka askeleella pitää ratkaista epälineaarinen algebrallinen yhtälö. Eksplisiittisessä tapauksessa taas $x_{n+\ell}$ saadaan suoraan aiemmin laskettujen arvojen avulla.

Toinen tapa parantaa tarkkuutta on ottaa enemmän ”näytteitä” suuntakentästä jokaista askelta kohti. Huomaa, että moniaskelmenetelmissä joka askeleella tarvitsee laskea f vain yhdessä uudessa pisteessä. Näissä *monivaihe*menetelmissä eli *Runge-Kutta menetelmissä* siis lasketaan f :n arvoja useissa pisteissä ennen kuin otetaan askel. Yleinen (eksplisiittinen) m -vaiheinen RK-menetelmä voidaan kirjoittaa muodossa

$$\begin{aligned}\kappa_1 &= f(t_n, x_n) \\ \kappa_2 &= f(t_n + c_2 h, x_n + h a_{21} \kappa_1) \\ &\vdots \\ \kappa_m &= f(t_n + c_m h, x_n + h(a_{m1} \kappa_1 + \cdots + a_{m,m-1} \kappa_{m-1})) \\ x_{n+1} &= x_n + h(b_1 \kappa_1 + \cdots + b_m \kappa_m)\end{aligned}\tag{6.4}$$

Luonnollisesti Eulerin menetelmä on tämänkin erikoistapaus: $m = b_1 = 1$; ja yleisesti ottaen yksivaiheiset RK-menetelmät on sama asia kuin yksiaskeliset moniaskelmenetelmät!

Teoriassa voisi ajatella, että voitaisiin muodostaa menetelmiä jotka ovat sekä moniaskelisiä että monivaiheisia.³ Näitä on itse asiassa tutkittu, mutta on osoittautunut, että näistä saatava hyöty on liian pieni menetelmän monimutkaisuudesta aiheutuvaan vaivaan nähden.

Sekä moniaskel- että RK-menetelmissä on siis vapaita parametreja jotka pitäisi valita siten, että saataisiin mahdollisimman ”hyvä” ja ”tarkka” menetelmä. Olisi siis hyödyllistä olla jokin kriteeri, kuten numeerisen integroinnin yhteydessä ollut kertaluku, jonka perusteella parametrit voisi valita. Osoittautuu, kuten integroinnin yhteydessä, että kertaluku ei ole riittävä kriteeri vaan pitää vaatia tietynlainen stabiilisuus. Moniaskelmenetelmien ja RK-menetelmien analyysi on varsin erilaista. Tarkastellaan ensin RK-menetelmiä.

6.4 Runge-Kutta-menetelmät

Numeerisessa integroinnissa kertaluku määriteltiin sen mukaan miten menetelmä integroi tarkasti polynomeja. Lähdetään tässäkin liikkeelle siitä, että voidaanko yksinkertaisten difyhtälöitten ratkaisut saada tarkasti. Yksinkertaisin mahdollinen tapaus on tietysti $x' = 0$. Tämän ratkaisuthan ovat vakioita, ja voisi ajatella, että järkevän menetelmän pitäisi ratkaista tarkasti näin yksinkertainen tehtävä. RK-menetelmillä tämä toteutuu automaattisesti. Menetelmän määrittävien kaavojen (6.4) viimeinen rivi voidaan kirjoittaa

$$x_{n+1} = x_n + hF(t_n, x_n, h)$$

missä F on yleisesti ottaen hyvin monimutkainen funktio. On kuitenkin selvää, että jos $f = 0$, niin myös $F = 0$, joten mikä tahansa RK-menetelmä antaa ratkaisuksi

³Näitä sanotaan joskus *yleisiksi lineaarisiksi menetelmiksi* (general linear methods).

$$x_n = \dots = x_0 = x(t_0).$$

Entäpä sitten tapaus $x' = 1$, $x(0) = 0$ jonka ratkaisu on $x(t) = t$. Tällöin selvästi $\kappa_j = 1$ kaikilla j joten tässä tapauksessa saadaan

$$x_{n+1} = x_n + hF(t_n, x_n, h) = x_n + h \sum_{j=1}^m b_j$$

Kun valitaan $x_0 = x(0) = 0$, niin selvästi $x_n = x(t_n) = nh$, jos $\sum_{j=1}^m b_j = 1$. Tämän pitemmälle ei oikein intuition perusteella päästä vaan tarvitaan seuraava

Määritelmä 6.3. Olkoon $x_n = x(t_n)$. Menetelmän (6.4) kertaluku on s , jos

$$x(t_n + h) - x_{n+1} = \mathcal{O}(h^{s+1})$$

Nyt voidaankin tarkistaa, että Eulerin menetelmän kertaluku on yksi. Taylorin Lauseen mukaan

$$x(t_n + h) = x(t_n) + hx'(t_n) + \mathcal{O}(h^2) = x(t_n) + hf(t_n, x_n) + \mathcal{O}(h^2)$$

Siis jos $x_n = x(t_n)$, niin $x(t_n + h) - x_{n+1} = \mathcal{O}(h^2)$.

Menetelmän kertaluvun määrittäminen yleisessäkin tapauksessa on *periaatteessa* yksinkertaista Taylorin kehitelmän avulla (käytännön laskut ovat sitten hyvinkin työläitä...). Otetaan käyttöön seuraavat lyhennysmerkinnät. Jos on selvää missä pisteessä f halutaan laskea niin merkitään $f = f(t, x)$ ja samoin asetetaan

$$f_t = \frac{\partial f}{\partial t}(t, x) \quad \text{ja} \quad f_x = \frac{\partial f}{\partial x}(t, x)$$

Esimerkiksi siis jos x on tehtävän $x' = f(t, x)$ ratkaisu niin

$$\frac{d}{dt}f(t, x(t)) = f_t + f_x x'(t) = f_t + f_x f$$

Esimerkki 6.3. Tarkastellaan 2-vaiheista menetelmää

$$\begin{aligned} \kappa_1 &= f(t_n, x_n) \\ \kappa_2 &= f(t_n + c_2 h, x_n + ha_{21}\kappa_1) \\ x_{n+1} &= x_n + h(b_1\kappa_1 + b_2\kappa_2) = x_n + hF(t_n, x_n, h) \end{aligned}$$

Haluttaisiin määrittää vakiot siten, että kertaluku on mahdollisimman iso. Pitäisi siis laskea Taylorin kehitelmän termejä sekä approksimaatiolle x_{n+1} että tarkalle ratkaisulle $x(t_n + h)$. Ensin saadaan

$$\begin{aligned} x(t_n + h) &= x(t_n) + hx'(t_n) + \frac{1}{2}h^2x''(t_n) + \mathcal{O}(h^3) \\ &= x(t_n) + hf(t_n, x_n) + \frac{1}{2}h^2(f_t + f_x f) + \mathcal{O}(h^3) \end{aligned}$$

Kun kehitetään x_{n+1} Taylorin sarjaksi niin tulkitaan F pelkästään h :n funktioksi. Pitäisi siis laskea

$$F(h) = F(0) + F'(0)h + \mathcal{O}(h^2)$$

Koska nyt oletetaan, että $x_n = x(t_n)$, niin $F(0) = (b_1 + b_2)f$ ja

$$F'(0) = c_2 f_t + a_{21} f_x f$$

Siispä

$$x_{n+1} = x_n + h F(h) = x(t_n) + h(b_1 + b_2)f + h^2 b_2 (c_2 f_t + a_{21} f_x f) + \mathcal{O}(h^3)$$

joten

$$x(t_n + h) - x_{n+1} = f(1 - b_1 - b_2)h + \frac{1}{2} \left((1 - 2b_2 c_2) f_t + (1 - 2b_2 a_{21}) f_x f \right) h^2 + \mathcal{O}(h^3)$$

Nähdään siis, että menetelmän kertaluku on ainakin yksi jos $b_1 + b_2 = 1$. Jotta saataisiin toisen kertaluvun menetelmä niin pitää vaatia, että $2b_2 c_2 = 2b_2 a_{21} = 1$. Selvästi siis $c_2 = a_{21}$, mutta edelleen on useita vaihtoehtoja valita parametrit. Yhtälöt toteutuvat, jos valitaan esimerkiksi $b_1 = b_2 = 1/2$ ja $a_{21} = 1$, jolloin siis

$$\begin{aligned} \kappa_1 &= f(t_n, x_n) \\ \kappa_2 &= f(t_n + h, x_n + h\kappa_1) \\ x_{n+1} &= x_n + \frac{h}{2} (\kappa_1 + \kappa_2) \end{aligned}$$

Tätä sanotaan joskus *Heunin menetelmäksi*. Voidaan osoittaa, että seuraavaa termiä ei enää saada nollaksi, joten kaksivaiheisen eksplisiittisen RK-menetelmän kertaluku on korkeintaan kaksi. ★

Edellisen esimerkin laskut yleistyvät seuraavasti.

Lemma 6.1. *m -vaiheisen RK-menetelmän kertaluku on vähintään yksi jos ja vain jos $\sum_{j=1}^m b_j = 1$. Lisäksi jos m -vaiheisen RK-menetelmän kertaluku on vähintään kaksi niin*

$$c_i = \sum_{j=1}^{i-1} a_{ij} \quad \text{kaikilla} \quad 2 \leq i \leq m$$

Todistus. Todistus on suora, joskin hiukan työläs, lasku joka etenee kuten edellisessä esimerkissä. ✧

Yhteenvedona voisi sanoa, että kaikki järkevät RK-menetelmät toteuttavat lemmas-
sa annetut ehdot. Kun sitten halutaan etsiä korkeamman kertaluvun menetelmiä niin laskut tulevat erittäin työläiksi ja näissä kannattaa sitten ottaa MAPLE tai jokin muu symbolisen laskennan ohjelmisto avuksi. Ennen tietokoneaikaa ehkäpä kaikkein käytetyin RK-menetelmä oli seuraava jota joskus sanotaan *klassiseksi RK-menetelmäksi*.

$$\begin{aligned} \kappa_1 &= f(t_n, x_n) \\ \kappa_2 &= f(t_n + h/2, x_n + h\kappa_1/2) \\ \kappa_3 &= f(t_n + h/2, x_n + h\kappa_2/2) \\ \kappa_4 &= f(t_n + h, x_n + h\kappa_3) \\ x_{n+1} &= x_n + \frac{h}{6} (\kappa_1 + 2\kappa_2 + 2\kappa_3 + \kappa_4) \end{aligned}$$

Tämän kertaluku on neljä. Huomaa erityisesti miten yksinkertaisia kertoimet ovat. Tarvittiin siis 4 vaihetta jotta saataisiin kertaluku 4. Valitettavasti korkeamman kertaluvun menetelmissä tarvitaan enemmän vaiheita.

m	5	6	7	8	9	10	11
max	4	5	6	6	7	7	8

Taulukko 6.1: Maksimikertaluku joka m -vaiheisella RK-menetelmällä voidaan saavuttaa.

Lemma 6.2. *Jos m -vaiheisen RK-menetelmän kertaluku on m , niin $m \leq 4$.*

Todistus löytyy kirjasta [10]. Taulukossa 6.1 on annettu maksimikertaluvut mitkä tietyllä vaihemäärällä voidaan saavuttaa. Tietääkseni korkein kertaluku mikä oikeasti on toteutettu tietokoneella on kahdeksan vaikka korkeammankin kertaluvun menetelmiä on olemassa.

Mutta miten kertaluvusta päästään itse numeerisen ratkaisun virheeseen? Tarkastellaan siis yleistä yhtälöä (6.1) ja halutaan numeerinen ratkaisu välille $[t_0, t_0 + T]$. Valitaan jokin N ja asetetaan $h = T/N$. Asetetaan edelleen alkuarvoksi $x_0 = x(t_0)$ ja lasketaan arvot x_n , $1 \leq n \leq N$ menetelmällä (6.4).

Lause 6.2. *Jos RK-menetelmän kertaluku on s , niin $\mathcal{E}_h = \mathcal{O}(h^s)$.*

Todistuksen idea. Jos halutaan ratkaisu välillä $[t_0, t_0 + T]$ niin pitää ottaa $N = T/h$ askelta. Kertaluvun määrittävä yhtälö voidaan tulkita yhden askeleen virheeksi. Jos siis joka askeleella tehdään virhe ε_n joka on luokkaa $\mathcal{O}(h^{s+1})$, niin voisi toivoa että kokonaisvirhe olisi $\mathcal{E}_h \approx N \sum_n \varepsilon_n = \mathcal{O}(h^s)$.

Tämä päättely onkin oikea. Tekninen vaikeus on osoittaa, että joka askeleella tehdyt pienet virheet eivät lähde kasvamaan nopeammin. Tarkka todistus löytyy esimerkiksi kirjasta [10]. $\diamond$

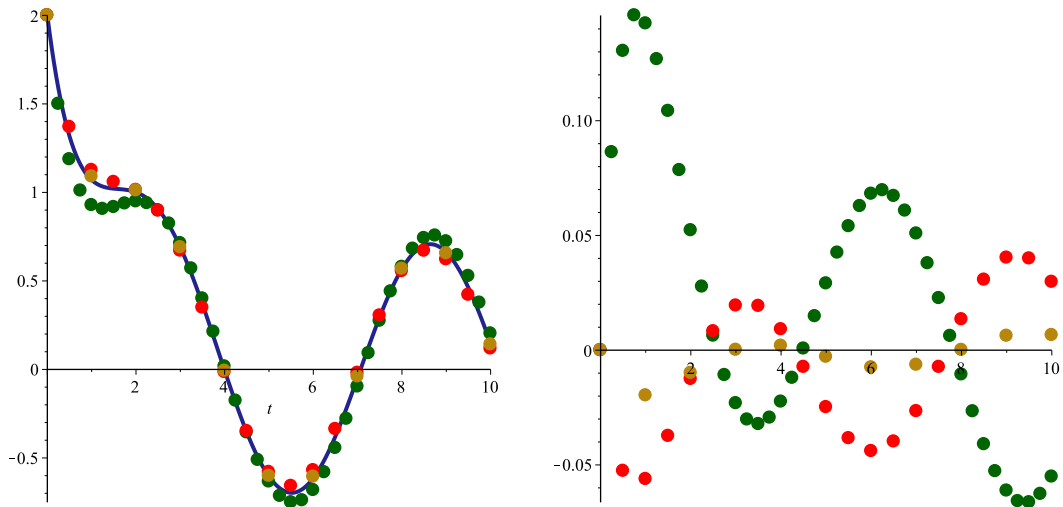
Siis jos RK-menetelmän kertaluku on vähintään yksi niin numeerinen ratkaisu supenee kohti tarkkaa ratkaisua kun $h \rightarrow 0$.

Esimerkki 6.4. Olkoon $x' = -x + \sin(t)$ ja valitaan alkuarvoksi $x(0) = 2$. Tarkka ratkaisu on $x(t) = \sin(t)/2 - \cos(t)/2 + 5e^{-t}/2$. Lasketaan numeerinen ratkaisu Eulerin menetelmällä askelpituudella 0.25, Heunin menetelmällä askelpituudella 0.5 ja klassisella RK-menetelmällä askelpituudella 1. Tällöin siis työmäärä kaikilla menetelmillä on sama siinä mielessä että pitää laskea yhtä monta funktion arvoa. Kuten kuvasta 6.4 nähdään niin korkeamman kertaluvun menetelmät ovat selvästi parempia jo näinkin isolla aika-askeleella. Kun halutaan tarkempi tulos ja siis pienennetään aika-askeleta, niin korkeamman kertaluvun menetelmien suhteellinen paremmuus kasvaa edelleen. $\star$

6.5 Moniaskelmenetelmät

Tarkastellaan yleistä moniaskelmenetelmää (6.3). Miten menetelmän kertoimet a_j ja b_j pitäisi valita? Katsotaan aluksi kaksiaskelista menetelmää, ja lähdetään liikkeelle integraalimuodosta

$$x(t_{n+2}) - x(t_{n+1}) = \int_{t_{n+1}}^{t_{n+2}} f(s, x(s)) ds$$



Kuva 6.4: Esimerkin 6.4 numeerinen ratkaisu verrattuna tarkkaan ratkaisuun vasemmalla ja virheet oikealla. Euler: vihreä, Heun: punainen, klassinen RK: kellanruskea.

Oletetaan, että ollaan jo laskettu pisteet $(t_0, x_0), \dots, (t_{n+1}, x_{n+1})$ ja merkitään $f_n = f(t_n, x_n)$ ja $g(t) = f(t, x(t))$. Nyt voidaan ajatella, että koska $x_n \approx x(t_n)$, niin

$$g(t) \approx \left(1 - \frac{t - t_n}{h}\right)g(t_n) + \frac{t - t_n}{h}g(t_{n+1}) \approx \left(1 - \frac{t - t_n}{h}\right)f_n + \frac{t - t_n}{h}f_{n+1}$$

Siispä

$$\int_{t_{n+1}}^{t_{n+2}} f(s, x(s))ds \approx \int_{t_{n+1}}^{t_{n+2}} \left(\left(1 - \frac{t - t_n}{h}\right)f_n + \frac{t - t_n}{h}f_{n+1} \right)ds = \frac{h}{2}(3f_{n+1} - f_n)$$

Päädytään siis seuraavaan menetelmään:

$$x_{n+2} - x_{n+1} = \frac{h}{2}(3f_{n+1} - f_n) \quad (6.5)$$

Vastaava implisiittinen menetelmä saadaan kun g :n interpoloinnissa käytetään myös pistettä (t_{n+2}, x_{n+2}) :

$$g(t) \approx p_2(t) = \frac{(t - t_{n+1})(t - t_{n+2})}{2h^2}f_n + \frac{(t - t_n)(t_{n+2} - t)}{h^2}f_{n+1} + \frac{(t - t_n)(t - t_{n+1})}{2h^2}f_{n+2}$$

Tästä integroimalla saadaan

$$\int_{t_{n+1}}^{t_{n+2}} f(s, x(s))ds \approx \int_{t_{n+1}}^{t_{n+2}} p_2(s)ds = \frac{h}{12}(5f_{n+2} + 8f_{n+1} - f_n)$$

mikä johtaa menetelmään

$$x_{n+2} - x_{n+1} = \frac{h}{12}(5f_{n+2} + 8f_{n+1} - f_n) \quad (6.6)$$

Esimerkki 6.5. Katsotaan taas yksinkertaista yhtälöä $x' = 1$, $x(0) = 0$ jonka ratkaisu on $x(t) = t$. Kaksiaskelisella menetelmällä pitäisi jostain sitten saada alkuarvo x_1 koska itse tehtävä antaa vain $x_0 = x(0) = 0$. Otetaan tässä tarkka arvo $x_1 = x(h) = h$ ja palataan tähän alkuarvo-ongelmaan myöhemmin. Soveltamalla menetelmää (6.5) saadaan $x_{n+2} - x_{n+1} = h$, joten $x_n = nh = x(t_n)$. Samoin (6.6) antaa tarkan tuloksen tässä tapauksessa. ★

Menetelmät (6.5) ja (6.6) näyttävät siis ainakin jossain määrin järkeviltä. Jotta päästäisiin analyysissä eteenpäin tarvitaan

Määritelmä 6.4. Olkoon $x_n = x(t_n), \dots, x_{n+\ell-1} = x(t_{n+\ell-1})$. Menetelmän (6.3) kertaluku on s , jos

$$x(t_{n+\ell}) - x_{n+\ell} = \mathcal{O}(h^{s+1})$$

Nyt näyttäisi siltä, että joudutaan taas laskemaan monimutkaisia Taylorin kehitelmiä kuten RK-menetelmien tapauksessa. Voidaan kuitenkin osoittaa, että moniaskelmenetelmien tapauksessa päästään paljon helpommalla. Moniaskelmenetelmien ominaisuuksia voidaan kätevästi analysoida tutkimalla menetelmän kertoimien määrittämiä polynomeja.

Määritelmä 6.5. Polynomeja

$$\rho(z) = \sum_{j=0}^{\ell} a_j z^j \quad \text{ja} \quad \sigma(z) = \sum_{j=0}^{\ell} b_j z^j$$

sanotaan menetelmän (6.3) karakteristisiksi polynomeiksi.

Esimerkiksi siis menetelmän (6.5) tapauksessa

$$\rho(z) = z^2 - z \quad \text{ja} \quad \sigma(z) = \frac{3}{2}z - \frac{1}{2}$$

Nyt voidaan osoittaa

Lemma 6.3. Menetelmän (6.3) kertaluku on s jos ja vain jos

$$\rho(e^h) - h\sigma(e^h) = \mathcal{O}(h^{s+1})$$

Todistus. Todistetaan vain, että annettu ehto on välttämätön. Riittävyys on todistettu kirjassa [10]. Sovelletaan menetelmää tehtävään $x' = x$, $x(0) = 1$ jonka ratkaisu on $x(t) = e^t$. Tällöin siis

$$\begin{aligned} \sum_{j=0}^{\ell} (a_j - hb_j)x_{n+j} &= (a_{n+\ell} - hb_{n+\ell})x_{n+\ell} + \sum_{j=0}^{\ell-1} (a_j - hb_j)e^{(n+j)h} = \\ &= (a_{n+\ell} - hb_{n+\ell})x_{n+\ell} + e^{nh} \left(\sum_{j=0}^{\ell-1} (a_j - hb_j)e^{hj} \right) = 0 \end{aligned}$$

Siispä

$$(a_{n+\ell} - hb_{n+\ell})(x(t_{n+\ell}) - x_{n+\ell}) = e^{nh} \left(\sum_{j=0}^{\ell} (a_j - hb_j) e^{hj} \right) = e^{nh} (\rho(e^h) - h\sigma(e^h))$$

Koska kertaluku on s , niin täytyy päteä

$$x(t_{n+\ell}) - x_{n+\ell} = \frac{e^{nh}}{a_{n+\ell} - hb_{n+\ell}} (\rho(e^h) - h\sigma(e^h)) = \mathcal{O}(h^{s+1})$$

Tämä on mahdollista vain jos $\rho(e^h) - h\sigma(e^h) = \mathcal{O}(h^{s+1})$. ◇

Kehitetään siis Taylorin sarjaksi

$$g(h) = \rho(e^h) - h\sigma(e^h) = g_0 + g_1h + g_2h^2 + \dots$$

jolloin kertaluku on s , jos $g_0 = g_1 = \dots = g_s = 0$. Heti nähdään, että $g_0 = g(0) = \rho(1)$ joten polynomilla ρ on aina nollakohta $z = 1$. Yleisesti ottaen voidaan osoittaa

Lemma 6.4. *Olkoon g kuten yllä. Tällöin*

$$g_0 = \rho(1) = \sum_{j=0}^{\ell} a_j \quad \text{ja}$$

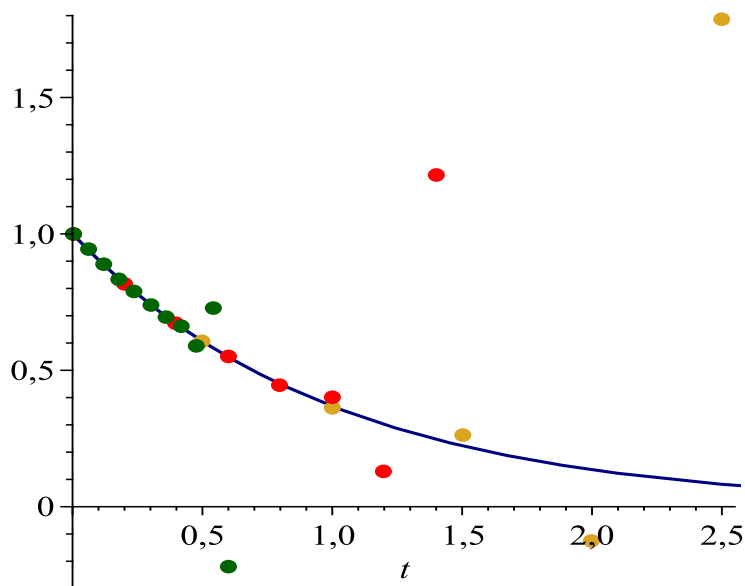
$$g_k = \frac{1}{k!} \sum_{j=0}^{\ell} a_j j^k - \frac{1}{(k-1)!} \sum_{j=0}^{\ell} b_j j^{k-1} \quad , \quad k \geq 1$$

Nyt on suoraviivaista tarkistaa, että menetelmän (6.5) kertaluku on kaksi ja menetelmän (6.6) kertaluku on kolme. Menetelmät (6.5) ja (6.6) ovat erikoistapauksia seuraavasta paljon käytetystä menetelmäperheestä.

Määritelmä 6.6. Adamsin menetelmissä *valitaan ensin $a_\ell = -a_{\ell-1} = 1$ ja $a_j = 0$, jos $j < \ell - 1$. Tämän jälkeen lasketaan kertoimet b_j siten, että kertaluku on ℓ eksplisiittisessä ja $\ell + 1$ implisiittisessä tapauksessa.*

Yksiaskelinen tapaus antaa jälleen tutut menetelmät: eksplisiittinen yksiaskelinen Adamsin menetelmä on Eulerin menetelmä ja implisiittinen yksiaskelinen Adamsin menetelmä on trapetsimenetelmä. Lemman 6.4 avulla voidaan osoittaa, että Adamsin menetelmissä kertoimet b_j voidaan valita siten, että saadaan määritelmässä annetut kertaluvut.

Adamsin menetelmien tapauksessa kertaluku on siis ℓ eksplisiittisessä ja $\ell + 1$ implisiittisessä tapauksessa. Koska ℓ -askelisessa menetelmässä on $2\ell + 1$ vapaata parametria implisiittisessä ja 2ℓ eksplisiittisessä tapauksessa niin voisi ajatella, että voitaisiin löytää paljon korkeamman kertaluvun menetelmiä. Tämä onkin mahdollista.



Kuva 6.5: Esimerkin 6.6 menetelmällä saadut ratkaisut kolmella eri askelpituudella.

Esimerkki 6.6. Olkoon $\ell = 2$ ja etsitään eksplisiittistä menetelmää. Normalisoinnin $a_2 = 1$ jälkeen on 4 vapaata parametria, joten voisi toivoa, että nämä voisi valita siten, että kertaluku olisi 3. Tämä johtaa yhtälöihin

$$\begin{aligned}\rho(1) &= g_0 = a_0 + a_1 + 1 = 0 \\ g_1 &= a_1 + 2 - b_0 - b_1 = 0 \\ g_2 &= \frac{1}{2}(a_1 + 4) - b_1 = 0 \\ g_3 &= \frac{1}{6}(a_1 + 8) - \frac{1}{2}b_1 = 0\end{aligned}$$

Tämän ratkaisuna saadaan kolmannen kertaluvun menetelmä

$$x_{n+2} + 4x_{n+1} - 5x_n = h(4f_{n+1} + 2f_n)$$

Sovelletaan tätä tehtävään $x' = -x$, $x(0) = 1$ jonka ratkaisu on $x(t) = e^{-t}$. Käytetään taas ”ylimääräisenä” alkuarvona tarkkaa arvoa $x_1 = e^{-h}$. Kuten kuvasta 6.5 nähdään, niin numeerinen ratkaisu on värähtelee yhä voimakkaammin, ja askeleen pienentäminen vain aiheuttaa sen, että epästabiilisuus ilmenee aiemmin. ★

Mistä sitten esimerkin menetelmän ongelmat johtuvat? Sovelletaan menetelmää yhtälöön $x' = 0$ jolloin saadaan differenssiyhtälö

$$x_{n+2} + 4x_{n+1} - 5x_n = 0$$

Koska $\rho(z) = z^2 + 4z - 5 = (z - 1)(z + 5)$ niin Lemman A.8 mukaan tämän yleinen ratkaisu on $x_n = c_1 + c_2(-5)^n$, missä vakiot c_1 ja c_2 riippuvat alkuehdoista x_0 ja x_1 . Jos nyt valitaan $x(0) = 1$ jolloin tarkka ratkaisu on $x(t) = 1$, niin tästä saadaan

$$x_n = \frac{5 + x_1}{6} + \frac{1 - x_1}{6}(-5)^n$$

Nyt ratkaisun ensimmäinen termi approksimoi haluttua vakioratkaisua, ja toinen termi on eräänlainen ”loisratkaisu”. Jos toinen alkuarvo $x_1 = 1$ niin loisratkaisu on nolla, mutta jos $x_1 \neq 1$, niin loisratkaisu kasvaa rajatta, ja hyvin nopeasti ratkaisu ei enää ole lähelläkään haluttua vakioratkaisua.

Äskeisessä esimerkissä käytettiin ”tarkkaa” arvoa $x_1 = e^{-h}$ mutta tästä huolimatta ratkaisu oli epästabiili. Tämä johtuu siitä, että kun lasketaan liukuluvuilla, niin koskaan ei voi olla varma täsmällisestä yhtäsuuruudesta. Iteroinnin kuluessa siis enemmän tai myöhemmin eksponentiaalisesti kasvavan termin kerroin muuttuu nolasta poikkeavaksi. Tässä siis ongelma johtui siitä, että ρ :lla oli itseisarvoltaan ykköistä suurempia nollakohtia. Näitä pitää siis välttää.

Määritelmä 6.7. *Moniaskelmenetelmä (6.3) on nollastabiili, jos differenssiyhtälön*

$$\sum_{j=0}^{\ell} a_j x_{n+j} = 0$$

ratkaisut pysyvät rajoitettuna mielivaltaisilla alkuehdoilla.

Lemman A.9 perusteella saadaan heti

Lemma 6.5. *Moniaskelmenetelmä (6.3) on nollastabiili, jos polynomin ρ nollakohdille z_j pätee: $|z_j| \leq 1$ kaikilla j ja ne nollakohdat joilla $|z_j| = 1$ ovat yksinkertaisia.*

Yleisesti ottaen ρ :lla on siis aina nollakohta $z = 1$. Vaikka teoriassa muut yksikkökehällä olevat nollakohdat eivät aiheuta stabiilisuuongelmia, niin käytännössä tällaisia menetelmiä halutaan välttää. Tämän takia asetetaan

Määritelmä 6.8. *Menetelmä (6.3) on vahvasti stabiili, jos $\rho(z) = (z - 1)\hat{\rho}(z)$ ja polynomin $\hat{\rho}$ nollakohdille z_j pätee: $|z_j| < 1$ kaikilla j .*

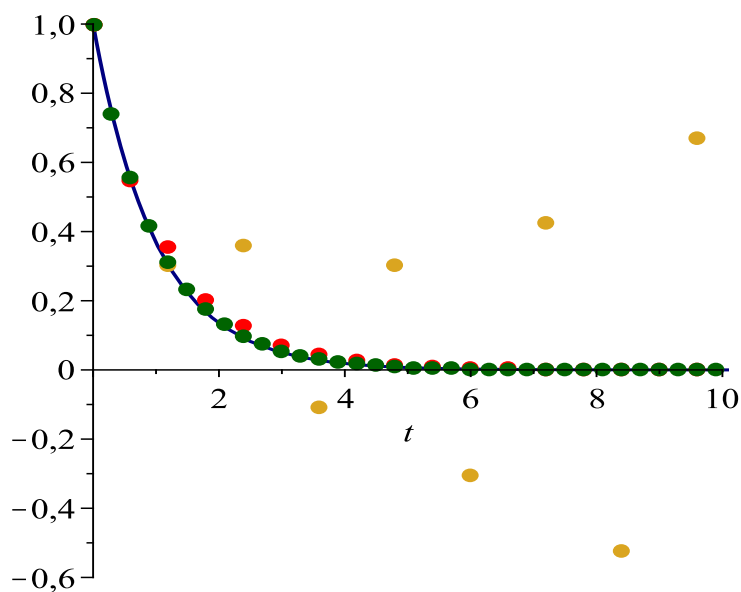
Tämän voisi tulkita niin, että nollastabiilisuus takaa sen, että loisratkaisut pysyvät rajoitettuna, kun $n \rightarrow \infty$. Sen sijaan jos menetelmä on vahvasti stabiili, niin loisratkaisut menevät nollaan kun $n \rightarrow \infty$. Adamsin menetelmät ovat aina vahvasti stabiileja koska tällöin $\rho(z) = (z - 1)z^{\ell-1}$.

Koska vain stabiileja menetelmiä voidaan käyttää numeerisissa laskuissa niin luonnollinen kysymys on kuinka korkea kertaluku stabiileilla menetelmillä voi olla. Tähän vastaa seuraava lause joka tunnetaan nimellä *Dahlquist's first barrier* (todistus löytyy kirjasta [10]).

Lause 6.3. *Jos ℓ -askelinen menetelmä (6.3) on vahvasti stabiili, niin sen kertaluku on korkeintaan ℓ , jos menetelmä on eksplisiittinen ja $\ell + 1$ jos menetelmä on implisiittinen.*

Stabiilisuusvaatimus pudottaa siis kertaluvun noin puoleen teoreettisesta maksimikertaluvusta. Toisaalta Adamsin menetelmät osoittavat, että tämä stabiilien menetelmien korkein mahdollinen kertaluku tosiaan voidaan saavuttaa.

Nyt voidaankin sitten osoittaa



Kuva 6.6: Menetelmällä (6.5) saatuja numeerisia ratkaisuja esimerkin 6.7 tapauksessa eri aika-askelilla.

Lause 6.4. Jos menetelmä (6.3) on nollastabiili ja sen kertaluku $s \geq 1$, niin numeerinen ratkaisu suppenee kohti tarkkaa ratkaisua ja

$$\mathcal{E}_h = \mathcal{O}(h^s)$$

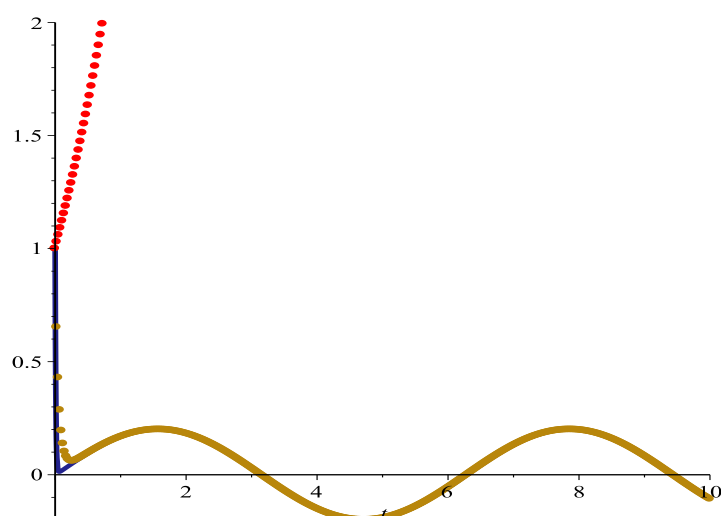
Todistuksen idea. Todistus on saman tyyppinen kuin RK-menetelmien tapauksessa. Yksi eksponentti ”menetetään” koska pitää ottaa $N = T/h$ askelta. Nollastabiilisuutta tarvitaan, etteivät yhden askeleen virheet rupea kasvamaan hallitsemattomasti. ✧

Esimerkki 6.7. Lause 6.4 takaa, että riittävän pienellä h numeerinen ratkaisu on lähellä tarkkaa ratkaisua. Isommilla h :n arvoilla numeerinen ratkaisu voi käyttäytyä hyvinkin eri tavalla kuin tarkka ratkaisu. Sovelletaan menetelmää (6.5) tehtävään $x' = -x$, $x(0) = 1$, jonka ratkaisu on $x(t) = e^{-t}$. Kuten kuvasta 6.6 nähdään niin isommilla h :n arvoilla numeerinen ratkaisu kasvaa rajatta. Kun aika-askelta sitten pienentää tarpeeksi saadaan odotettu tulos. ★

6.6 Erityisiä tehtävaluokkia ja muita kommentteja

6.6.1 Kankeat tehtävät

Edellä on jo saatu (riittävän) korkean kertaluvun RK-menetelmiä ja moniaskelmenetelmiä. Miksi siis tarvittaisiin enää muita menetelmiä?



Kuva 6.7: Esimerkin 6.8 tehtävän ratkaisu klassisella RK-menetelmällä kahdella eri aika-askeleella.

Esimerkki 6.8. Tarkastellaan ongelmaa

$$\begin{cases} x' = -100x + 20 \sin(t) \\ x(0) = 1 \end{cases}$$

Kuvassa 6.7 on ratkaistu tämä tehtävä klassisella RK-menetelmällä kun $h = 1/400$ (kellanuskea) ja $h = 1/357$ (punainen). Alussa ratkaisulla on nopea transientti, jolloin askeleen pitäisikin olla pieni. Sitten ratkaisu asettuu hitaasti muuttuvalle jaksolliselle radalle, jolloin askel voisi tarkkuuden puolesta olla huomattavasti isompi. Tämä on kuitenkin mahdotonta koska numeerinen ratkaisu alkaisi kasvaa eksponentiaalisesti.

★

Tämäntyyppisiä tehtäviä sanotaan *kankeiksi* (*stiff*). Kankeudelle ei ole löytynyt mitään matemaattisen tarkkaa määritelmää, mutta seuraava käytännön huomio on ainakin oikean suuntainen.

”Määritelmä”

Tehtävä on kankea, jos eksplisiittisen menetelmän täytyy käyttää tarpeettoman pieniä askeleita.

Toisin sanoen mikään eksplisiittinen menetelmä ei voi olla hyvä kankeille tehtäville vaan tarvitaan välttämättä implisiittisiä menetelmiä. On osoittautunut, että seuraava menetelmäperhe soveltuu hyvin kankeille tehtäville.

Määritelmä 6.9. BDF-menetelmissä asetetaan ensin $b_\ell = 1$ ja $b_j = 0$ jos $j < \ell$. Tämän jälkeen lasketaan kertoimet a_j siten, että kertaluku on ℓ .

Lemman 6.4 perusteella voidaan osoittaa, että kertoimet a_j tosiaan voidaan valita siten, että saadaan haluttu kertaluku. Ensimmäinen BDF-menetelmä on

$$x_{n+1} - x_n = h f(t_{n+1}, x_{n+1})$$

jota luonnollisesti sanotaan *implisiittiseksi Eulerin menetelmäksi*. BDF-menetelmiä sanotaan joskus *Gearin menetelmiksi* koska Gear oli ensimmäinen joka huomasi näitten menetelmien käytännön merkityksen [8]. Huomattakoon, että implisiittisyys sinänsä ei vielä takaa mitään; esimerkiksi implisiittiset Adamsin menetelmät eivät sovellu kankeille tehtäville.

Lemma 6.6. *BDF-menetelmät ovat vahvasti stabiileja, kun $\ell \leq 6$.*

Todistus. Tämä on suora lasku. ✧

Voidaan osoittaa, että BDF-menetelmät ovat epästabiileja kun $\ell > 6$, mutta tällä ei ole juurikaan käytännön merkitystä koska kertaluku 6 on täysin riittävä. BDF-menetelmät soveltuvat kankeille tehtäville koska niillä on vahvan stabiilisuuden lisäksi eräs toinen stabiilisuusominaisuus. Tarkastellaan jälleen testiongelmää $x' = -ax$, $x(0) = 1$ missä $a > 0$. Tämän ratkaisu $x(t) = e^{-at}$ lähestyy nollaa kun $t \rightarrow \infty$. Nyt voisi toivoa, että numeerinen menetelmä tuottaisi jonon x_n joka myös lähestyisi nollaa kun $n \rightarrow \infty$. Esimerkiksi Eulerin menetelmällä tämä ei aina toteudu.

$$x_{n+1} = x_n + hf_n = (1 - ah)x_n \Rightarrow x_n = (1 - ah)^n x_0$$

Siis $x_n \rightarrow 0$ vain jos $h < 1/a$. BDF-menetelmille ei tule tällaista rajoitusehtoa aika-askeleelle. Tarkistetaan tämä seuraavan kaksiaskeleisen BDF-menetelmän tapauksessa.

$$3x_{n+2} - 4x_{n+1} + x_n = 2hf_{n+2}$$

Lemma 6.7. *Testiongelmään sovellettuna BDF2-menetelmän ratkaisulle pätee $x_n \rightarrow 0$ kun $n \rightarrow \infty$ kaikilla $a > 0$ ja kaikilla $h > 0$.*

Todistus. Pitää siis analysoida seuraavan differenssiyhtälön ratkaisuja.

$$(3 + 2ah)x_{n+2} - 4x_{n+1} + x_n = 0$$

Tämän karakteristinen polynomi on $p(z) = (3 + 2ah)z^2 - 4z + 1$ jonka nollakohdat ovat

$$z_{\pm} = \frac{2 \pm \sqrt{1 - 2ah}}{3 + 2ah}$$

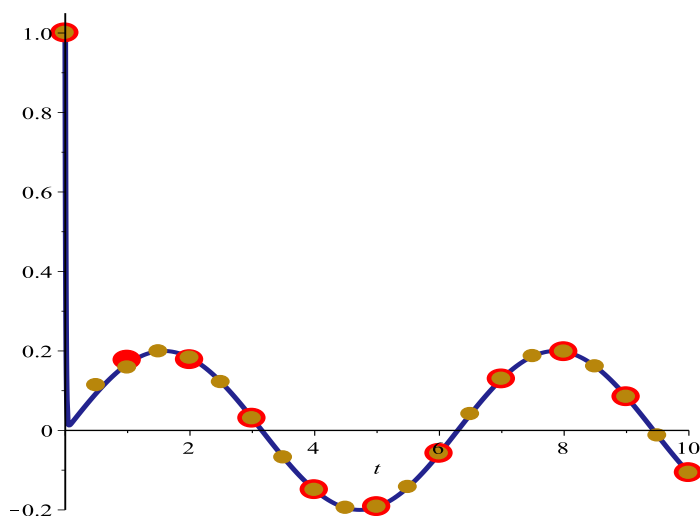
Lemman A.8 mukaan yleinen ratkaisu on

$$x_n = c_1 z_+^n + c_2 z_-^n$$

Selvästi nyt $|z_{\pm}| < 1$ kaikilla $h > 0$ joten $x_n \rightarrow 0$ kun $n \rightarrow \infty$. ✧

Esimerkki 6.9. Sovelletaan BDF2-menetelmää esimerkin 6.8 tehtävään. Kuvasta 6.8 nähdään, että jo hyvinkin isolla aika-askeleella saadaan hyvä tulos. ★

Kankeitten tehtävien tapauksessa BDF-menetelmillä voidaan siis ottaa huomattavasti isompia askeleita kuin "tavallisilla" menetelmillä. Tästä saatava hyöty on paljon suurempi kuin implisiittisyydestä aiheutuva ylimääräinen vaiva. Kankeita tehtäviä on perusteellisesti analysoitu kirjassa [11].



Kuva 6.8: Esimerkin 6.8 tehtävän ratkaisu BDF2-menetelmällä kahdella eri aika-askeleella.

6.6.2 Difyhtälösystemit

Useimmiten luonnollisesti halutaan ratkaista difyhtälösystemejä. Onneksi tähän ei tarvita mitään uutta teoriaa, vaan yllä käsitellyt menetelmät soveltuvat sellaisinaan tapaukseen jossa sekä x että f ovat vektoreita.

6.6.3 Käytännön toteutus

Käytännössä haluttaisiin muuttaa automaattisesti askelpituutta laskennan kuluessa. Tämä tapahtuu siten, että tavalla tai toisella yritetään arvioida virhettä kun otetaan uusi askel. Jos arvioidaan, että virhe on liian iso, niin askel hylätään, pienennetään askelta, ja yritetään uudestaan. Koska menetelmä suppenee, niin riittävän pienellä askeleella päästään lopulta eteenpäin.

Toisaalta jos arvioidaan, että yhden askeleen virhe on erittäin pieni, niin tällöin tietenkin hyväksytään askel, mutta myös kasvatetaan seuraavaa askelta jotta laskenta olisi mahdollisimman tehokasta. Voisi siis sanoa, että pyrkimyksenä on, että jokaisella (hyväksytyllä) askeleella tehty virhe on suurin piirtein vakio riippumatta askelpituudesta. Kirjassa [10] on selvitetty miten tämä käytännössä voidaan toteuttaa.

Moniaskelmenetelmien tapauksessa voidaan joskus vaihtaa paitsi askelta, myös menetelmän kertalukua. Toisin sanoen on ohjelmoitu kaikki Adamsin tai BDF-menetelmät esimerkiksi kertalukuun 6 asti, ja ohjelma saattaa laskennan kuluessa tietyissä olosuhteissa siirtyä kertaluvusta toiseen.

Implisiittisiä menetelmiä tarvittiin ennen kaikkea kankeitten tehtävien takia. Tällöin ongelmaksi muodostuu miten ratkaista yhtälö

$$a_\ell x_{n+\ell} - h b_\ell f(t_{n+\ell}, x_{n+\ell}) = v$$

missä v on tunnettu. Merkitään $y = x_{n+\ell}$. Ehkäpä yksinkertaisin idea olisi käyttää kiintopisteiteraatiota

$$a_\ell y_{k+1} = h b_\ell f(t_{n+\ell}, y_k) + v$$

Huomaa, että nyt on varsin helppoa valita hyvä alkuarvo, esimerkiksi $y_0 = x_{n+\ell-1}$. Selvästi tämä iteraatio suppenee kun h on riittävän pieni. Muistetaan kuitenkin, että implisiittisiä menetelmiä käytetään nimenomaan kankeille tehtäville. Nyt osoittautuu, että käytännössä jos tehtävä on kankea, niin kiintopisteiteraatiossa pitäisi valita tarpeettoman pieni h , jolloin siis menetetään se etu jota implisiittisellä menetelmällä haettiin. Siis käytännössä epälineaarinen systeemi ratkaistaan Newtonin menetelmällä.

6.6.4 Hamiltonin systeemit

Eräs tärkeä tehtäväluokka joka vaatii erityisiä numeerisia menetelmiä ovat *Hamiltonin systeemit*. Näitä esiintyy paljon klassisen mekaniikan ja monikappaledynamiikan ongelmissa siinä tapauksessa että systeemin kitka voidaan jättää huomiotta, jolloin systeemin kokonaisenergia säilyy.

Tarkastellaan seuraavassa vain yhtä esimerkkiä jossa nähdään millaisia ongelmia tästä numeerisille menetelmille aiheutuu. Tarkastellaan heilurisysteemiä

$$\begin{cases} \theta' = v \\ v' = -\sin(\theta) \end{cases} \quad (6.7)$$

missä θ on heilahduskulma ja v on nopeus. Nyt voidaan helposti tarkistaa, että

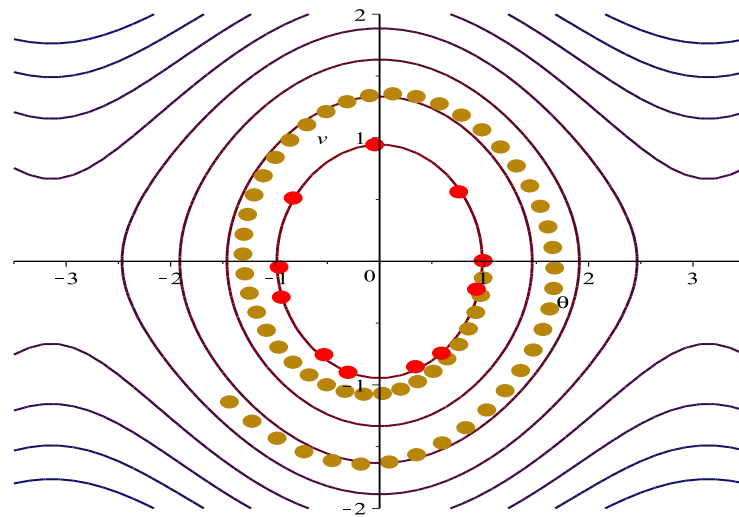
$$H = \frac{1}{2} v^2 - \cos(\theta)$$

jota sanotaan tehtävän *Hamiltonin funktioksi*, pysyy vakiona ratkaisua pitkin. Siis ratkaisun radat (θ, v) tasossa ovat H :n tasa-arvokäyriä. Kuvassa 6.9 on ratkaistu tehtävä sekä Eulerin menetelmällä että klassisella RK-menetelmällä. Lähellä origoa tarkat ratkaisut, siis H :n tasa-arvokäyrät ovat umpinaisia, joten ratkaisut ovat jaksollisia. Eulerin menetelmä ei kuitenkaan yhden kierroksen jälkeen ”osu” lähellekään jaksollista ratkaisua vaikka aika-askel on suhteellisen pieni. Klassinen RK-menetelmää pärjää luonnollisesti huomattavasti paremmin, mutta senkin kanssa tulee ongelmia kun laskenta-aikaa kasvatetaan.

Ongelmat johtuvat pohjimmiltaan siitä, että Hamiltonin systeemeillä on tietty erityisrakenne, ja vain eräät numeeriset menetelmät ottavat sopivalla tavalla tämän rakenteen huomioon. Näitä Hamiltonin systeemeille sopivia menetelmiä sanotaan *symplektisiksi menetelmiksi*. Niitä on analysoitu kirjassa [9].

6.7 Harjoitustehtäviä

1. Laske 3-askelinen Adamsin eksplisiittinen ja implisiittinen menetelmä. Tarkista kertaluvut.
2. Laske 3-askelinen BDF menetelmä. Tarkista, että kertaluku on 3 ja että menetelmä on vahvasti stabiili.



Kuva 6.9: Tehtävän (6.7) ratkaisu sekä Eulerin menetelmällä että klassisella RK-menetelmällä.