

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Közönséges differenciálegyenletek numerikus megoldása

Szakdolgozat

Mészáros Mirjana

Matematika BSc - Matematikai elemző szakirány

Témavezető: Kurics Tamás

Alkalmazott Analízis és Számításmatematikai Tanszék



2010

Köszönöm témavezetőmnek, Kurics Tamásnak,
hogy a tanórák keretében és konzultációk alkalmával is
segített megismerkedni ezzel a témával, továbbá,
hogy hasznos tanácsaival hozzájárult a szakdolgozat megírásához.

Tartalomjegyzék

1. Motiváció	2
2. Bevezetés	3
3. Egylépéses módszerek	4
3.1. Az Euler-módszer	4
3.1.1. A módszer leírása	4
3.1.2. Az Euler-módszer rendje	6
3.2. A javított Euler-módszer	8
3.2.1. A módszer leírása	8
3.2.2. A javított Euler-módszer rendje	8
3.3. Implicit Euler-módszer	9
3.3.1. A módszer leírása	9
3.3.2. Az implicit Euler-módszer rendje	9
3.4. Trapéz módszer	10
3.4.1. A trapéz módszer leírása	10
3.4.2. A trapéz módszer rendje	10
3.5. A θ -módszer	11
3.6. Egylépéses módszerek tesztelése	11
4. Runge-Kutta típusú módszerek	15
4.1. Az explicit Runge-Kutta módszer	15
4.2. A kétlépcsős Runge-Kutta módszerek	16
4.3. Harmadrendű módszerek	18
4.4. Negyed- és magasabbrendű módszerek	19
4.5. Runge-Kutta módszerek tesztelése	19
5. Többlépéses módszerek	22
5.1. A középpont szabály	22
5.2. Az Adams-Bashforth és Adams-Moulton módszerek	23
5.2.1. Kétlépéses Adams-Bashforth módszer	24
5.2.2. Kétlépéses Adams-Moulton módszer	25
5.3. A Nyström és a Milne-Simpson módszerek	26
5.4. A prediktor-korrektor módszerek	26
5.5. Többlépéses módszerek tesztelése	27
6. Összefoglalás	32

1. Motiváció

A mindennapi életben találkozunk olyan megoldandó problémákkal, ahol a probléma modellezése differenciálegyenletekkel történik. Ezek lehetnek gazdasági, műszaki vagy akár biológiai problémák.

Vegyük az alábbi példát, amikor egy hír terjedését vizsgáljuk egy közösségben (lásd [2]). Kezdetben a hírt a teljes lakosság y sűrűségű része ismeri, $1 - y$ nem ismeri. Ahhoz, hogy a hír továbbterjedjen, a két csoportnak találkoznia kell, ennek a valószínűsége $y(1 - y)$. A találkozás sikeressége egy α pozitív konstanssal jellemezhető. Arra számítunk, hogy ahogy az idő (t) múlik, annál nagyobb lesz a hírt ismerők aránya az egész lakosságban. Ha a t időpillanatban $y(t)$ volt a sűrűség, akkor a találkozás után, a $t + \Delta t$ időpontban

$$y(t + \Delta t) = y(t) + \Delta t \alpha y(t)(1 - y(t)),$$

azaz

$$\frac{y(t + \Delta t) - y(t)}{\Delta t} = \alpha y(t)(1 - y(t)).$$

Ebből adódik a következő differenciálegyenlet:

$$y'(t) = \alpha y(t)(1 - y(t)).$$

Egy másik érdekes példa egy fajon belül a populációszám vizsgálata (lásd [1]). Kezdjük a vizsgálatot a t_0 időben, és legyen ekkor az egyedszám P_0 . Arra vagyunk kíváncsiak, hogy mennyi lesz az egyedszám egy $t = t_1$ pillanatban. Az egyedszámot két esemény befolyásolhatja, a születések és a halálozások. Adódhatnak még más befolyásoló tényezők, mint például a ki- és bevándorlás, betegségek, járványok, háborúk, de ezekkel most nem foglalkozunk. Legyen b a Δt idő alatt születettek aránya a teljes lakossághoz képest, hasonlóan c a halálozás aránya. A modell felírva így néz ki:

$$P(t + \Delta t) = P(t) + bP(t)\Delta t - cP(t)\Delta t,$$

vagy másképp

$$\frac{P(t + \Delta t) - P(t)}{\Delta t} = bP(t) - cP(t) = kP(t).$$

Ebből, az előbbi példához hasonlóan, kapjuk a

$$P'(t) = kP(t)$$

differenciálegyenletet.

A példaként felhozott differenciálegyenletek megoldhatóak analitikusan is, viszont a gyakorlatban felmerülő problémák gyakran nemlineáris egyenletre vagy egyenletrendszerre vezetnek, ilyenkor numerikus módszerekhez fordulhatunk, ezeknek a módszereknek az összefoglalásával és tanulmányozásával foglalkozik a szakdolgozatom.

2. Bevezetés

Feladatunk az

$$u'(t) = f(t, u(t)) \quad (1)$$

alakú közönséges differenciálegyenlet megoldása, ahol $f(t, u)$ adott függvény és keressük u -t. Az (1) egyenlet magában nem elég ahhoz, hogy egyértelműen meghatározzuk az u függvényt. Nézzük a legegyszerűbb esetet, amikor $f(t, u)$ azonosan 0. Ekkor u valamilyen konstans függvény, hiszen egy függvény deriváltja pontosan akkor nulla, ha a függvény konstans. Hogy mennyi ez a konstans, azt egyértelműen nem tudjuk meghatározni, további információra van szükségünk. Ezt a plusz információt hívjuk kezdeti értéknek, azaz megmondjuk, hogy mennyi a függvény értéke a $t = 0$ -ban¹, azaz $u(0)$. Ezzel együtt megkaptuk a kezdetiérték feladatot:

$$u'(t) = f(t, u(t)), \quad u(0) = u_0. \quad (2)$$

A megoldás létezését és egyértelműségét az alábbi tétel garantálja:

2.1. Picard-Lindelöf tétel. *Egy kezdetiérték feladatnak létezik megoldása, és a megoldás egyértelmű, ha:*

- f folytonos az első argumentumában, azaz t -ben
- f Lipschitz²-folytonos u -ban, azaz létezik olyan L állandó, hogy minden u_1, u_2 és minden t értékre teljesül az

$$|f(t, u_1) - f(t, u_2)| \leq L|u_1 - u_2| \quad (3)$$

egyenlőtlenség.

Miután definiáljuk a különböző numerikus módszereket, azok működését egy konkrét példán is tesztelni fogjuk, amit Matlab-ban programozok le. A tesztegyenletünk a következő lesz:

$$u'(t) = -\lambda u(t) + \lambda(t+1)^3 + 3(t+1)^2, \quad u(0) = u_0 \quad (4)$$

aminek ismerjük a pontos megoldását,

$$u(t) = t^3 + 3t^2 + 3t + 1 + (u_0 - 1)e^{-\lambda t}. \quad (5)$$

A λ paraméter és a τ lépésköz változtatásával teszteljük a numerikus módszereket a $[0, 4]$ intervallumban, a 4-beli értékét pedig összehasonlítjuk a pontos megoldás 4-beli értékével.

¹abban az esetben, ha a függvényt egy $[0, b]$ intervallumon vizsgáljuk

²Rudolf Otto Sigismund Lipschitz (1832-1903), német matematikus

3. Egylépéses módszerek

Az $u'(t) = f(t, u(t))$ megoldásához definiáljunk egy y_τ rácsfüggvényt, úgy, hogy a t tengelyen vizsgált intervallumot fölosztjuk N egyenlő részre, egy ilyen kis távolság τ hosszú lesz. A függvény értelmezési tartománya ezeknek a pontoknak a halmaza lesz, tehát:

$$\mathcal{D}(y_\tau) = \omega_\tau := \{n \cdot \tau; n = 0, 1, \dots, N; \tau > 0\}$$

ekvidisztáns rácsháló, $t_n = n \cdot \tau$. Néhány egyszerűsítő jelölést is bevezetünk, legyen

$$y_n := y_\tau(t_n),$$

ez a közelítő megoldás értéke a t_n pontban,

$$u_n := u(t_n)$$

pedig a pontos megoldás a t_n pontban. A rácsfüggvény, y_τ pontos definíciója az egyes módszereknél fog szerepelni.

3.1. Az Euler-módszer

3.1.1. A módszer leírása

A közönséges differenciálegyenletek numerikus megoldásának legegyszerűbb eszköze az Euler³-módszer, habár ezt csak ritkán használják, mert mint később látni fogjuk, elég pontatlan.

Ennél a módszernél a y_τ rácsfüggvényt az alábbi módon definiáljuk:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n) \quad (6)$$

$$y_0 = u_0 \quad (7)$$

Ezt átrendezve kapjuk, hogy:

$$y_{n+1} = y_n + \tau \cdot f(t_n, y_n). \quad (8)$$

Ezt máshogy úgy is megkaphatjuk, hogy abból indulunk ki, hogy ismerjük az $u'(t) = f(t, u(t))$ kezdetiérték feladatot az u_0 értékkel együtt, így ismerjük az $u(t)$ deriváltját a $t_0 = 0$ pontban: $u'(0) = f(0, u(0))$, és legyen ebben a kezdeti pontban $y'(0) = u'(0)$. Választunk egy kis $\tau > 0$ értéket, és a kiszámított derivált irányában teszünk egy τ vetületű lépést:

$$t_1 = t_0 + \tau = \tau, \quad y_1 = y_0 + \tau f(0, y_0). \quad (9)$$

³Leonhard Paul Euler (1707-1783), svájci matematikus

Ezek után újra kapunk egy $(t, y(t))$ pontot a síkon, melyben, az előzőekhez hasonlóan, a differenciálegyenletből kiszámolható a derivált, és így újra közelíthetjük a megoldást egy újabb τ hosszú szakaszon:

$$t_2 = t_1 + \tau, \quad y_2 = y_1 + \tau f(t_1, y_1). \quad (10)$$

Általánosan leírva:

$$t_{n+1} = t_n + \tau, \quad y_{n+1} = y_n + \tau f(t_n, y_n). \quad (11)$$

Nézzük meg a módszer lépéseit egy konkrét példán (lásd [2]):

$$u'(t) = Lu(t), \quad t \in [0, 1], \quad u(0) = 1, \quad (12)$$

ahol L pozitív konstans. Ennek tudjuk a pontos megoldását, $u(t) = e^{Lt}$, és így össze tudjuk hasonlítani a pontos megoldást a numerikus megoldással. Legyen $\tau = 0,1$ és $L = 10$.

t_n	$u(t_n)$	$y(t_n)$
0,1	2,71828	2
0,2	7,38906	4
0,3	20,0855	8
$\vdots$	$\vdots$	$\vdots$
1,0	22026,5	1024

Látszik, hogy az $y(t_n)$ eredmények mennyire távol vannak a pontos $u(t_n)$ értékektől. Folytathatjuk a számolást úgy, hogy egyre kisebb τ -t veszünk, $\tau = 1/N$, és mindig duplázzuk az intervallumok számát. $e_N := |u(t_N) - y(t_N)|$, vagyis a pontos megoldástól való eltérés, $t_N = 1 = N\tau$, y_N pedig a becsült megoldás az t_N pontban.

N	τ	y_N	e_N
10	0,1	1024	21002,5
20	0,05	3325,26	18701,24
40	0,025	7523,16	14503,34
80	0,0125	12365,2	9661,3
160	0,00625	16316,6	5709,9
320	0,003125	18900,3	3126,0
640	0,0015625	20387,5	1639,0

A pontos megoldás értéke $u(t_N) = 22026,5$. Látható, hogy ha lassan is, de az értékek konvergálnak a pontos megoldáshoz.

3.1.2. Az Euler-módszer rendje

3.1. Definíció. Egy adott numerikus módszer által előállított y_τ rácsfüggvény esetén, az $e(t_n, \tau)$ rácsfüggvényt a módszer hibafüggvényének vagy globális hibájának nevezzük,

$$e_n := y(t_n) - u(t_n) = y_n - u_n$$

$$e : \omega_\tau \rightarrow \mathbb{R} \quad (13)$$

Az Euler-módszer:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n). \quad (14)$$

Mivel $y_n = u_n + e_n$, ezért:

$$\frac{e_{n+1} - e_n}{\tau} = f(t_n, u_n + e_n) - \frac{u_{n+1} - u_n}{\tau}$$

$$\frac{e_{n+1} - e_n}{\tau} = [f(t_n, u_n + e_n) - f(t_n, u_n)] + [f(t_n, u_n) - \frac{u_{n+1} - u_n}{\tau}]. \quad (15)$$

$$\psi_n^{(1)} := -\frac{u_{n+1} - u_n}{\tau} + f(t_n, u_n) \quad (16)$$

$$\psi_n^{(2)} := f(t_n, u_n + e_n) - f(t_n, u_n) \quad (17)$$

A $\psi_n^{(1)}$ értéket nevezzük reziduális, vagy lokális approximációs hibának. $\psi_n^{(1)}$ nem más, mint a pontos megoldás értéke a numerikus megoldást előállító algoritmusban. Tehát ha $y_\tau = u(t)$, akkor $\psi_n^{(1)} = 0$.

3.2. Definíció. Azt mondjuk, hogy egy numerikus módszer konzisztens, ha a képlethibájára igaz, hogy $\psi_n^{(1)} = O(\tau^p)$, ahol $p > 0$, a legnagyobb ilyen egész p a módszer konzisztencia rendje.

Az Euler-módszer konzisztencia rendje ezek alapján:

$$\psi_n^{(1)} := -\frac{u_{n+1} - u_n}{\tau} + f(t_n, u_n)$$

$$\psi_n^{(1)} = -\frac{u(t_{n+1}) - u(t_n)}{\tau} + f(t_n, u(t_n)) \quad (18)$$

$$u(t_{n+1}) = u(t_n + \tau) = u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2} u''(t_n) + O(\tau^3). \quad (19)$$

A fenti felbontás az $u(t_{n+1})$ függvény t_n pontbeli Taylor-polinommal⁴ közelített alakja. A Taylor-polinom általános alakja:

$$T_n(x) = f(a) + f'(a)(x - a) + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n,$$

⁴Brook Taylor (1685-1731), angol matematikus

ha az f függvény n -szer differenciálható az a pontban.

Mivel $u(t)$ a kezdeti érték feladat megoldása, ezért $u'(t) = f(t, u(t))$ minden t -re.

$$\begin{aligned} u'(t_n) &= f(t_n, u(t_n)) \\ \psi_n^{(1)} &= -u'(t_n) - \frac{\tau}{2}u''(t_n) + O(\tau^2) + u'(t_n) \\ \psi_n^{(1)} &= -\frac{\tau}{2}u''(t_n) + O(\tau^2) = O(\tau^1) \end{aligned} \quad (20)$$

Tehát a módszer elsőrendben konzisztens.

Vonjuk ki a **3.1.** definícióban definiált egyenletből a pontos megoldást behelyettesítve a képletbe, azaz a

$$u_{n+1} = u_n + \tau f(t_n, u_n)$$

egyenletet, így a globális hibára az alábbi egyenleteket kapjuk:

$$e_{n+1} = e_n + \tau[f(t_n, y_n) - f(t_n, u_n)] + \psi_n^{(1)}\tau,$$

felhasználva a Lipschitz-tulajdonságot, következik a

$$|e_{n+1}| \leq |e_n| + L_f\tau|e_n| + \tau|\psi_n^{(1)}|$$

becslés, ami tovább fejtve

$$\begin{aligned} |e_{n+1}| &\leq (1 + L_f\tau)^2|e_{n-1}| + \tau|\psi_n^{(1)}| + (1 + L_f\tau)\tau|\psi_{n-1}^{(1)}| \\ &\leq \dots \leq (1 + L_f\tau)^{n+1}|e_0| + \sum_{k=0}^n (1 + L_f\tau)^{n-k}\tau|\psi_k^{(1)}| \\ &\leq (1 + L_f\tau)^{n+1} \left[|e_0| + \sum_{k=0}^n \tau|\psi_k^{(1)}| \right]. \end{aligned}$$

Itt

$$1 + L_f\tau \leq e^{L_f\tau},$$

tehát

$$(1 + L_f\tau)^n \leq e^{L_f\tau n} = e^{L_f t_n}.$$

Ezek után a végleges becslésünk a következő:

$$|e_{n+1}| \leq e^{L_f t_n} \left[|e_0| + \sum_{k=0}^n |\psi_k^{(1)}|\tau \right]. \quad (21)$$

3.3. Definíció. Ha $f \in Lip(y)$, akkor egy numerikus módszer stabil, ha igaz a következő:

$$|e_{n+1}| \leq C \left(|e_0| + \sum_{k=0}^n |\psi_k^{(1)}|\tau \right),$$

ahol C konstans.

Az Euler-módszer a fenti számolások miatt stabil.

3.1. Tétel. *Ha egy numerikus módszer konzisztens és stabil, akkor konvergens is, és a konvergencia rendje egybeesik a konzisztencia rendjével.*

Az Euler-módszer tehát mivel konzisztens és stabil is, ezért konvergens, a konvergencia rendje 1.

A további egylépéses módszereknél szintén érvényes lesz a stabilitás (ezt nem számoljuk ki a továbbiakban), ezért elég lesz csak a konzisztencia rendjét kiszámolni, amiből adódik majd a konvergencia rendje.

3.2. A javított Euler-módszer

A fenti számolások mutatják, hogy az Euler-módszer lassan konvergál, ezért ritkán is használják. Egy kis javítással, viszont jobb eredményeket várhatunk.

3.2.1. A módszer leírása

A módszert javíthatjuk úgy, hogy először egy fél lépést téve kiszámítjuk f értékét, és az így kapott meredekséggel teszünk egy egész lépést az eredeti (t_n, y_n) pontból. Azt várjuk, hogy az így definiált módszer másodrendű pontosságú lesz.

Képletekkel felírva:

$$t_{n+\frac{1}{2}} = t_n + \frac{\tau}{2}, \quad y_{n+\frac{1}{2}} = y_n + \frac{\tau}{2}f(t_n, y_n) \quad (22)$$

és az egész lépés az új meredekséggel:

$$t_{n+1} = t_n + \tau, \quad y_{n+1} = y_n + \tau f(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}). \quad (23)$$

3.2.2. A javított Euler-módszer rendje

A képlethiba kiszámolásához fel kell tennünk, hogy f mindkét változója szerint kétszer folytonosan differenciálható, így y háromszor folytonosan differenciálható lesz t szerint.

$$\begin{aligned} \frac{y_{n+1} - y_n}{\tau} &= f(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) \\ \psi_n^{(1)} &= -\frac{u_{n+1} - u_n}{\tau} + f(t_{n+\frac{1}{2}}, u_{n+\frac{1}{2}}) \end{aligned} \quad (24)$$

Az u_{n+1} és $f(t_{n+\frac{1}{2}}, u_{n+\frac{1}{2}}) = u'(t_{n+\frac{1}{2}})$ függvényeket Taylor-sorba fejtvé kapjuk, hogy:

$$\begin{aligned} \psi_n^{(1)} &= -\frac{1}{\tau}(u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2}u''(t_n) + O(\tau^3) - u(t_n)) \\ &\quad + u'(t_n) + \frac{\tau}{2}u''(t_n) + \frac{\tau^2}{8}u'''(t_n) + O(\tau^3) \end{aligned}$$

$$\begin{aligned}
\psi_n^{(1)} &= -u'(t_n) - \frac{\tau}{2}u''(t_n) + O(\tau^3) + u'(t_n) + \frac{\tau}{2}u''(t_n) + \frac{\tau^2}{8}u'''(t_n) + O(\tau^3) \\
&= \frac{\tau^2}{8}u'''(t_n) + O(\tau^3) = O(\tau^2).
\end{aligned} \tag{25}$$

Tehát a javított Euler-módszer tényleg másodrendű.

3.3. Implicit Euler-módszer

3.3.1. A módszer leírása

Az $u' = f(t, u(t))$ egyenlet közelítő megoldása az implicit Euler-módszerrel leírva így néz ki:

$$y_{n+1} = y_n + \tau f(t_{n+1}, y_{n+1}). \tag{26}$$

Ahogy látható a keresett y_{n+1} érték az egyenlet jobb oldalán is előfordul, ezért nevezzük ezt a módszert implicitnek. Ettől a módszertől azért várunk jobb eredményt, mert egy pontban a közelítő megoldást (az explicit Eulerrel ellentétben) az adott pontban (és nem az előzőben) mért meredekség segítségével közelíti.

3.3.2. Az implicit Euler-módszer rendje

Az explicit Euler-módszerhez hasonlóan számolhatjuk ki az implicit módszer rendjét. Ismét e -vel jelölve a hibafüggvényt, és a $e_n = y_n - u_n$ egyenletet használva:

$$\begin{aligned}
\frac{e_{n+1} - e_n}{\tau} &= f(t_{n+1}, u_{n+1} + e_{n+1}) - \frac{u_{n+1} - u_n}{\tau} \\
&= \left[f(t_{n+1}, u_{n+1} + e_{n+1}) - f(t_{n+1}, u_{n+1}) \right] + \left[f(t_{n+1}, u_{n+1}) - \frac{u_{n+1} - u_n}{\tau} \right]
\end{aligned}$$

Ebből:

$$\psi_n^{(1)} = -\frac{u_{n+1} - u_n}{\tau} + f(t_{n+1}, u_{n+1}). \tag{27}$$

Mivel $u(t)$ a kezdeti érték feladat megoldása, ezért $u'(t) = f(t, u(t))$ minden t -re.

$$u'(t_{n+1}) = f(t_{n+1}, u_{n+1})$$

Az $u(t_{n+1})$ kifejezést a már ismertek szerint felírhatjuk eképp:

$$u(t_{n+1}) = u(t_n + \tau) = u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2}u''(t_n) + O(\tau^3).$$

Kiszámolva ezek után $\psi_n^{(1)}$ értékét kapjuk, hogy

$$\psi_n^{(1)} = -u'(t_n) - \frac{\tau}{2}u''(t_n) + O(\tau^2) + u'(t_{n+1}).$$

Mivel

$$u'(t_{n+1}) = u'(t_n) + \tau u''(t_n) + O(\tau^2),$$

ezért

$$\psi_n^{(1)} = \frac{\tau}{2}u''(t_n) + O(\tau^2) = O(\tau^1). \quad (28)$$

A várttal ellentétben az implicit Euler-módszer, az explicithez hasonlóan, elsőrendű.

3.4. Trapéz módszer

3.4.1. A trapéz módszer leírása

Az explicit és implicit Euler-módszert elemezve láttuk, hogy a két módszer nagyon lassan konvergál az eredeti megoldáshoz. Az explicit módszernél, egy lépésköznél mindig csak az intervallum kezdőpontjában számoltuk ki a derivált értékét, és ez adta meg a folytatáshoz a meredekséget, míg az implicit módszernél csak az intervallum végpontjában számoltuk ki ezt az értéket. A trapéz módszernél az az ötlet, hogy vegyük az intervallum két végpontjában számolt értékek számtani közepét. Képlettel megfogalmazva:

$$y_{n+1} = y_n + \frac{1}{2}\tau[f(t_n, y_n) + f(t_{n+1}, y_{n+1})]. \quad (29)$$

3.4.2. A trapéz módszer rendje

A trapéz módszerre kiszámított ψ_n^1 lokális approxiációs hiba:

$$\psi_n^1 = -\frac{u(t_{n+1}) - u(t_n)}{\tau} + \frac{1}{2}[f(t_n, u(t_n)) + f(t_{n+1}, u(t_{n+1}))]. \quad (30)$$

Ezek után kiszámolhatjuk $u(t_{n+1})$ Taylor-polinomos alakját.

$$\begin{aligned} u(t_{n+1}) - u(t_n) &= u(t_n + \tau) - u(t_n) = u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2}u''(t_n) + O(\tau^3) - u(t_n) \\ \frac{u(t_{n+1}) - u(t_n)}{\tau} &= u'(t_n) + \frac{\tau}{2}u''(t_n) + O(\tau^2) \end{aligned} \quad (31)$$

Mivel

$$f(t_n, u(t_n)) = u'(t_n)$$

és

$$f(t_{n+1}, u(t_{n+1})) = u'(t_{n+1}) = u'(t_n) + \tau u''(t_n) + O(\tau^2)$$

az előzőleg látott Taylor-polinomos felbontás miatt, a ψ_n^1 kiszámolva:

$$\begin{aligned} \psi_n^1 &= -u'(t_n) - \frac{\tau}{2}u''(t_n) + O(\tau^2) \\ &+ \frac{1}{2}u'(t_n) + \frac{1}{2}u'(t_n) + \frac{1}{2}\tau u''(t_n) + O(\tau^2) = O(\tau^2). \end{aligned} \quad (32)$$

A trapéz módszer tehát a fenti számolások miatt másodrendű.

3.5. A θ -módszer

Az eddig megvizsgált módszereknél a közelítő megoldást a (t_{n+1}, y_{n+1}) pontban úgy kaptuk, hogy vettük az y_n és y_{n+1} pontbeli deriváltak valamilyen átlagolását. Ezt összefoglalva, felírhatjuk őket általános alakban is, és általánosan vizsgálhatjuk a módszer rendjét is. Ezt nevezzük θ -módszernek, ahol a θ az átlagolás súlyparamétere, azaz formálisan:

$$y_{n+1} = y_n + \tau[\theta f(t_n, y_n) + (1 - \theta)f(t_{n+1}, y_{n+1})], \quad (33)$$

$\theta \in [0, 1]$, aminek különböző értékeire különböző numerikus módszereket kapunk. Ha

- $\theta = 1$, az explicit Euler-módszert, ha
- $\theta = \frac{1}{2}$, a trapéz módszert, ha
- $\theta = 0$, az implicit Euler-módszert kapjuk.

A numerikus módszer képletébe beírva a pontos megoldást, felírva $u(t_{n+1})$ -et Taylor-sorba fejtvé és kihasználva, hogy $f(t_n, u(t_n)) = u'(t_n)$, megmondhatjuk a módszerünk rendjét.

$$\begin{aligned} \psi_n^{(1)} &= \frac{u(t_{n+1}) - u(t_n)}{\tau} - [\theta f(t_n, u(t_n)) + (1 - \theta)f(t_{n+1}, u(t_{n+1}))] \\ &= \frac{u(t_{n+1}) - u(t_n)}{\tau} - [\theta u'(t_n) + (1 - \theta)u'(t_{n+1})] \\ &= \frac{1}{\tau} \left[u(t_n) + \tau u'(t_n) + \frac{1}{2} \tau^2 u''(t_n) + \frac{1}{6} \tau^3 u'''(t_n) - u(t_n) \right] \\ &\quad - \left\{ \theta u'(t_n) + (1 - \theta) \left[u'(t_n) + \tau u''(t_n) + \frac{1}{2} \tau^2 u'''(t_n) \right] \right\} + O(\tau^3) \\ &= \left(\theta - \frac{1}{2} \right) \tau u''(t_n) + \left(\frac{1}{2} \theta - \frac{1}{3} \right) \tau^2 u'''(t_n) + O(\tau^3). \end{aligned} \quad (34)$$

Tehát a módszer másodrendű, ha $\theta = \frac{1}{2}$ (trapéz módszer), különben meg elsőrendű (explicit és implicit Euler-módszer).

3.6. Egylépéses módszerek tesztelése

Miután több egylépéses módszerrel megismerkedtünk és kiszámoltuk a konvergencia rendjüket, a bevezetőben említettek szerint, egy konkrét példán teszteljük is őket. A teszt-egyenletünk a következő:

$$u'(t) = -\lambda u(t) + \lambda(t+1)^3 + 3(t+1)^2, \quad u(0) = u_0$$

aminek a pontos megoldása

$$u(t) = t^3 + 3t^2 + 3t + 1 + (u_0 - 1)e^{-\lambda t}.$$

A program futtatása során a λ paramétert és a τ lépésközt fogom változtatni, az u_0 kezdetiérték minden esetben 10 lesz, a numerikus módszereket a $[0, 4]$ intervallumban vizsgálom és a végpontjában, 4-ben, megnézem, hogy a pontos és a közelített megoldás mennyivel tér el.

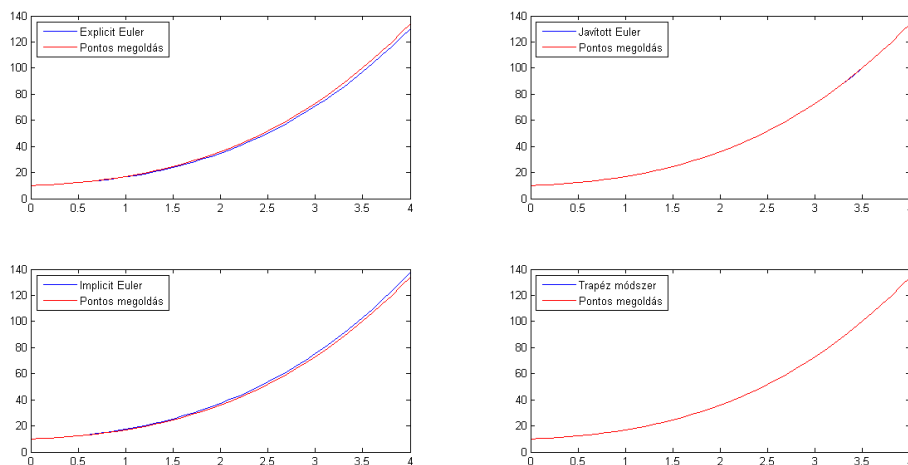
A τ lépésközt először 0,1-nek választom, majd megfelezve 0,05 és 0,025-nek, míg mindhárom lépésköznél a λ három értékével, 0, 15 és 30-cal, futtatom a programot.

1. táblázat. Egylépéses módszerek hibája különböző τ -val és λ -val

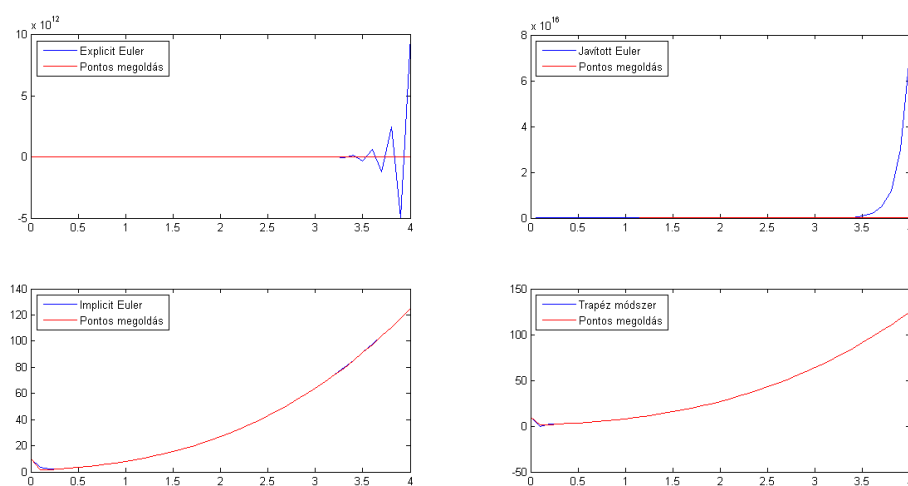
		Explicit Euler	Javított Euler	Implicit Euler	Trapéz módszer
$\tau = 0,1$	$\lambda = 0$	3,5800	0,0100	3,6200	0,0200
	$\lambda = 15$	0,0993	0,1418	0,0980	$3,33 \cdot 10^{-4}$
	$\lambda = 30$	$9,91 \cdot 10^{12}$	$7,46 \cdot 10^{16}$	0,0493	$1,67 \cdot 10^{-4}$
$\tau = 0,05$	$\lambda = 0$	1,7950	0,0025	1,8050	0,0050
	$\lambda = 15$	0,0495	0,0146	0,0492	$8,33 \cdot 10^{-5}$
	$\lambda = 30$	0,0249	0,0365	0,0247	$4,17 \cdot 10^{-5}$
$\tau = 0,025$	$\lambda = 0$	0,8988	$6,25 \cdot 10^{-4}$	0,9013	$1,20 \cdot 10^{-3}$
	$\lambda = 15$	0,0247	$2,80 \cdot 10^{-3}$	0,0246	$2,08 \cdot 10^{-5}$
	$\lambda = 30$	0,0124	$3,70 \cdot 10^{-3}$	0,0124	$1,04 \cdot 10^{-5}$

A 1. táblázatból látszik, hogy az elsőrendű explicit és implicit Euler-módszer szinte mindig rosszabb eredményt ad mind a két másodrendű módszer, a javított Euler és a trapéz-módszer (lásd 1., 3., 4. ábra). Az egyetlen kivétel, amikor a $\tau = 0,1$ és $\lambda = 30$ (lásd 2. ábra), ugyanis ilyenkor egy merev differenciálegyenletet kell megoldanunk, és arra az implicit módszerek sokkal jobb közelítést adnak. A másik érdekesség ami megfigyelhető, hogy ugyanolyan λ mellett ha a lépéstávot felezzük, az elsőrendű módszerek rendje feleződik, a másodrendűeké negyedelődik.

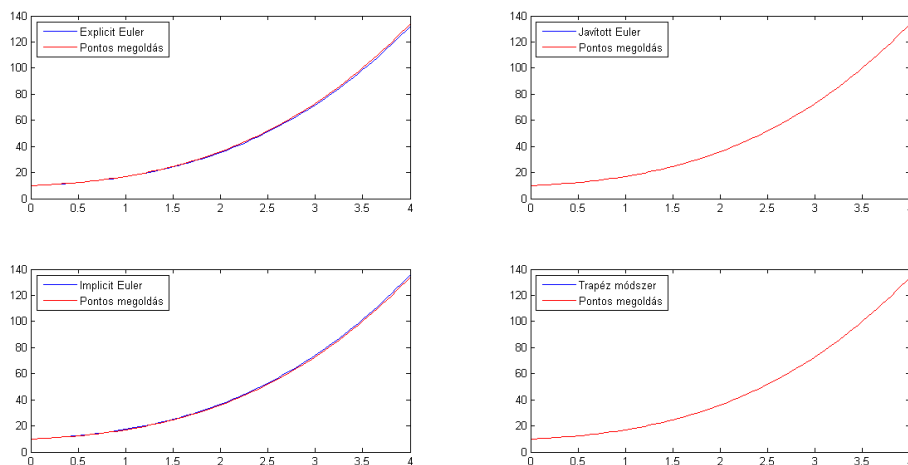
1. ábra. Egylépéses módszerek, $\lambda = 0$, $\tau = 0,1$



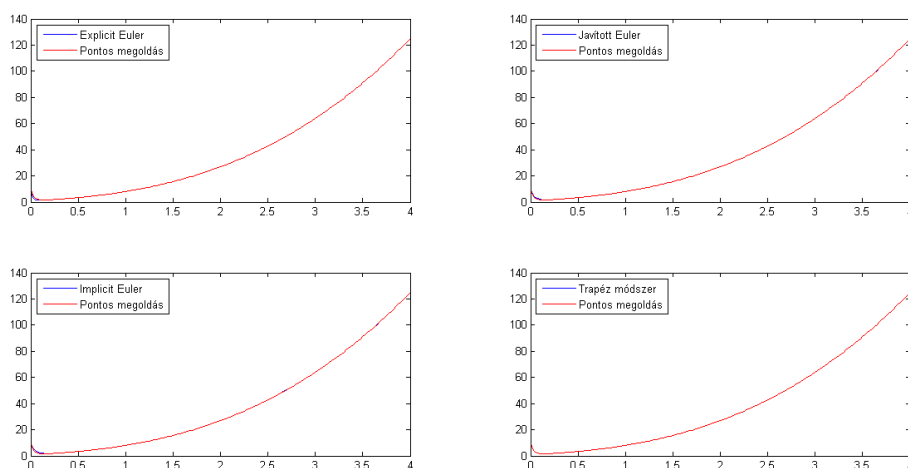
2. ábra. Egylépéses módszerek, $\lambda = 30$, $\tau = 0,1$



3. ábra. Egylépéses módszerek, $\lambda = 0$, $\tau = 0,05$



4. ábra. Egylépéses módszerek, $\lambda = 30$, $\tau = 0,025$



4. Runge-Kutta típusú módszerek

4.1. Az explicit Runge-Kutta módszer

Ebben a fejezetben megismerkedünk a Runge-Kutta⁵ módszerekkel, amik szintén egy-lépésesek, viszont az eddig megismert módszerekkel ellentétben az a jellemzőjük, hogy a numerikus eredmény kiszámolásához több értéket, "lépcsőt", kell kiszámolnunk, így majd magasabbrendű módszereket is elő tudunk állítani.

Definiáljuk rekurzívan az alábbi ξ számokat:

$$\begin{aligned}\xi_1 &= y_n \\ \xi_2 &= y_n + \tau a_{2,1} f(t_n, \xi_1) \\ \xi_3 &= y_n + \tau a_{3,1} f(t_n, \xi_1) + \tau a_{3,2} f(t_n + c_2 \tau, \xi_2) \\ &\vdots \\ \xi_m &= y_n + \tau \sum_{i=1}^{m-1} a_{m,i} f(t_n + c_i \tau, \xi_i),\end{aligned}$$

és ezek lineáris kombinációjából kapjuk meg az y_{n+1} értéket:

$$y_{n+1} = y_n + \tau \sum_{j=1}^m b_j f(t_n + c_j \tau, \xi_j). \quad (35)$$

Az m szám jelöli, hogy a Runge-Kutta módszer hány lépcsőből áll.

Az $a_{i,j}$ elemekből létrehozhatunk egy $m \times m$ méretű $\mathbf{A}$ mátrixot, ahol az üres helyeket nullával töltjük föl,

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ a_{2,1} & 0 & 0 & \dots & 0 \\ a_{3,1} & a_{3,2} & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & \\ a_{m,1} & a_{m,2} & \dots & a_{m,m-1} & 0 \end{pmatrix}$$

a b_i és c_i értékeket pedig vektorba rendezhetjük.

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_m \end{pmatrix}$$

⁵Carl David Tolmé Runge (1856-1927), Martin Wilhelm Kutta (1867-1944), német matematikusok

Az A mátrix és a b és c vektorok egyértelműen meghatározzák, hogy melyik Runge-Kutta módszerről beszélünk, ezért ezeket az elemeket szokás egy táblázatba rendezni, amit Butcher⁶-táblázatnak hívunk.

A Butcher-táblát a következő képpen kell kitölteni:

0					
c_2	$a_{2,1}$				
c_3	$a_{3,1}$	$a_{3,2}$			
$\vdots$	$\vdots$		$\ddots$		
c_m	$a_{m,1}$	$a_{m,2}$	$\dots$	$a_{m,m-1}$	
	b_1	b_2	$\dots$	b_{m-1}	b_m

Az egylépéses módszerekhez hasonlóan, ezeknél a módszereknél is csak a konzisztenciát fogjuk vizsgálni, a stabilitást feltesszük, így megkapjuk majd a konvergencia rendjét. A Runge-Kutta módszerek stabilitását az alábbi tétel garantálja.

4.1. Tétel. *Legyen $f \in Lip(y)$ L_f Lipschitz-állandóval, ekkor a Runge-Kutta módszer stabil, azaz*

$$|e_{n+1}| \leq e^{L_\Phi} \left(|e_0| + \sum_{k=0}^n |\psi_k^{(1)}| \tau \right),$$

ahol

$$L_\Phi = L_f \left(\sum_{j=1}^s |b_j| + Q_{s-1}(\tau L_f) \right),$$

és a $Q_{s-1}(\tau L_f)$ tag a τL_f -nek $(s-1)$ -edfokú polinomja, amely eleget tesz a $Q_{s-1}(\tau L_f) = O(\tau L_f)$ egyenlőségnek.

4.2. A kétlépcsős Runge-Kutta módszerek

Először a kétlépcsős módszereket vizsgáljuk és arra vagyunk kíváncsiak, hogy létezik-e kétlépcsős harmadrendű módszer. Először a $f(t_n + c_2\tau, \xi_2)$ kifejezésnek írjuk le a Taylor polinomos kifejtését a (t_n, y_n) pont körül.

$$\begin{aligned} f(t_n + c_2\tau, \xi_2) &= f(t_n + c_2\tau, y_n + a_{2,1}\tau f(t_n, y_n)) \\ &= f(t_n, y_n) + \tau \left[c_2 \frac{\partial f(t_n, y_n)}{\partial t} + a_{2,1} \frac{\partial f(t_n, y_n)}{\partial y} f(t_n, y_n) \right] + O(\tau^2) \end{aligned}$$

A (35) egyenletből

$$\begin{aligned} y_{n+1} &= y_n + \tau(b_1 + b_2)f(t_n, y_n) \\ &+ \tau^2 b_2 \left[c_2 \frac{\partial f(t_n, y_n)}{\partial t} + a_{2,1} \frac{\partial f(t_n, y_n)}{\partial y} f(t_n, y_n) \right] + O(\tau^3) \end{aligned} \quad (36)$$

⁶John C. Butcher (1933-), Új-zélandi matematikus

adódik.

Ezek után kifejtethetjük a pontos megoldást szintén a (t_n, y_n) körül. Jelöljük a pontos megoldást a t_{n+1} -ben $\tilde{y}$ -nal. Ekkor

$$\tilde{y}(t_{n+1}) = y_n + \tau f(t_n, y_n) + \frac{1}{2}\tau^2 \left[\frac{\partial f(t_n, y_n)}{\partial t} + \frac{\partial f(t_n, y_n)}{\partial y} f(t_n, y_n) \right] + O(\tau^3). \quad (37)$$

Összehasonlítva a (36) és a (37) egyenleteket, azt kapjuk, hogy

$$b_1 + b_2 = 1, \quad b_2 c_2 = \frac{1}{2}, \quad a_{2,1} = c_2.$$

Ezek a feltételek sok kombinációra igazak, viszont a legnépszerűbbek a következők:

$$\begin{array}{c|cc} 0 & & \\ \hline \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array}, \quad \begin{array}{c|cc} 0 & & \\ \hline \frac{2}{3} & \frac{2}{3} & \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}, \quad \begin{array}{c|cc} 0 & & \\ \hline 1 & 1 & \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}. \quad (38)$$

A rend vizsgálatához bevezetünk néhány egyszerűsítő jelölést. Legyen

$$b := b_2, \quad \text{ekkor} \quad b_1 = 1 - b$$

$$c := c_2 = a_{2,1}.$$

Így (35)-ből kapjuk, hogy

$$\begin{aligned} \frac{y_{n+1} - y_n}{\tau} &= (1 - b)f(t_n, y_n) + bf(t_n + c\tau, y_n + c\tau f(t_n, y_n)) \\ &= y_n - by_n + by_n + bc\tau y_n. \end{aligned}$$

Tehát

$$\frac{y_{n+1} - y_n}{\tau} = (1 + bc\tau)y_n. \quad (39)$$

Kiszámolva végül $\psi_n^{(1)}$ -t

$$\begin{aligned} \psi_n^{(1)} &= -\frac{u(t_{n+1}) - u(t_n)}{\tau} + (1 + bc\tau)u(t_n) \\ &= -u'(t_n) - u''(t_n)\frac{\tau}{2} - u'''(t_n)\frac{\tau^2}{6} + O(\tau^3) + (1 + bc\tau)u(t_n) \quad ^7 \\ &= -u(t_n) \left[\tau \left(\frac{1}{2} - bc \right) + \frac{\tau^2}{6} \right] + O(\tau^3) \quad ^8 \\ &= -u(t_n)\frac{\tau^2}{6} + O(\tau^3) = O(\tau^2), \end{aligned} \quad (40)$$

látjuk, hogy a τ^2 tag nem tűnik el, tehát nem lehet másodrendűnél jobb kétlépcsős módszert csinálni.

⁷ $u(t_{n+1})$ -t kifejtettük a Taylor-polinomjával

⁸ $u'(t) = u(t) \rightarrow u'''(t) = u''(t) = u'(t) = u(t)$

4.3. Harmadrendű módszerek

Ha harmadrendű módszert szeretnénk előállítani, akkor háromlépcsős Runge-Kutta módszert kell választanunk, ennél már a ξ_3 értéke is szerepelni fog a módszert megadó képletben.

$$\begin{aligned}\xi_1 &= y_n, \\ \xi_2 &= y_n + \tau a_{2,1} f(t_n, \xi_1), \\ \xi_3 &= y_n + \tau a_{3,1} f(t_n, \xi_1) + \tau a_{3,2} f(t_n + c_2 \tau, \xi_2),\end{aligned}$$

és

$$y_{n+1} = y_n + \tau \left(b_1 f(t_n, y_n) + b_2 f(t_n + c_2 \tau, \xi_2) + b_3 f(t_n + c_3 \tau, \xi_3) \right). \quad (41)$$

Ha ξ_2 és ξ_3 értékeket Taylor-sorba fejtenénk és így visszahelyettesítenénk a képletbe, azt kapnánk, hogy a módszer harmadrendű, és az ismeretlen $a_{i,j}, b_i, c_i$ értékekre is kapnánk megszorító feltételeket, amik a következők:

$$\begin{aligned}b_1 + b_2 + b_3 &= 1, & b_2 a_{2,1} + b_3 c_3 &= \frac{1}{2}, \\ c_2 &= a_{2,1}, & c_3 &= a_{3,1} + a_{3,2}, \\ \frac{1}{2} b_2 a_{2,1}^2 + \frac{1}{2} b_3 c_3^2 &= \frac{1}{6}, & b_3 a_{3,2} a_{2,1} &= \frac{1}{6}.\end{aligned} \quad (42)$$

Ezeknek a kritériumoknak egy speciális esete amikor

$$b_1 = b_3 = \frac{1}{6}, b_2 = \frac{2}{3}, c_3 = a_{3,1} + a_{3,2} = 1,$$

ezt Kutta harmadrendű módszerének nevezzük. A további ismeretleneket kiszámolva:

$$\begin{aligned}c_2 &= a_{2,1} = \frac{1}{2} \\ a_{3,2} &= 2, & a_{3,1} &= -1.\end{aligned}$$

Tehát a következő háromlépcsős módszer harmadrendben konvergál:

$$\begin{aligned}\xi_1 &= f(t_n, y_n) \\ \xi_2 &= y_n + \frac{\tau}{2} f(t_n, \xi_1) \\ \xi_3 &= y_n - \tau f(t_n, y_n) + 2\tau f\left(t_n + \frac{\tau}{2}, \xi_2\right) \\ y_{n+1} &= y_n + \frac{\tau}{6} f(t_n, y_n) + \frac{2\tau}{3} f\left(t_n + \frac{\tau}{2}, \xi_2\right) + \frac{\tau}{6} f(t_n + \tau, \xi_3).\end{aligned} \quad (43)$$

4.4. Negyed- és magasabbrendű módszerek

Negyedrendű módszert csak néglépcsős Runge-Kuttával kaphatunk, ekkor 10 ismeretlent tartalmazó egyenletrendszert kell megoldanunk, melynek több megoldása is lehet. A legismertebb néglépcsős módszer az úgynevezett klasszikus Runge-Kutta módszer:

$$\begin{array}{c|ccc}
 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & & \\
 \frac{1}{2} & 0 & \frac{1}{2} & \\
 1 & 0 & 0 & 1 \\
 \hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
 \end{array} \quad (44)$$

Ha magasabbrendű numerikus módszert próbálunk meg előállítani, például ötödrendűt, azt vesszük észre, hogy nem elég annyi lépcsőt használni, ahányrendű módszert szeretnénk kapni, például az ötödrendűhöz hatlépcsős Runge-Kutta módszer kell. Az alábbi összefüggések érvényesek a rend (p) és a lépcsőszám (m) között:

$1 \leq m \leq 4:$	$p \leq m$
$m \geq 5$	$p \leq m - 1$
$m \geq 8$	$p \leq m - 2$
$m \geq 10$	$p \leq m - 3$

(45)

4.5. Runge-Kutta módszerek tesztelése

A Runge-Kutta módszerek közül négyet fogunk tesztelni, két kétlépcsős módszert,

$$\begin{array}{c|cc}
 0 & & \\
 1 & 1 & \\
 \hline
 & \frac{1}{2} & \frac{1}{2}
 \end{array}, \quad
 \begin{array}{c|cc}
 0 & & \\
 \frac{2}{3} & \frac{2}{3} & \\
 \hline
 & \frac{1}{4} & \frac{3}{4}
 \end{array}, \quad (46)$$

egy háromlépcsőset

$$\begin{array}{c|ccc}
 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & & \\
 1 & -1 & 2 & \\
 \hline
 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6}
 \end{array}, \quad (47)$$

valamint egy néglépcsőset, a klasszikus Runge-Kutta módszert

$$\begin{array}{c|ccc}
 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & & \\
 \frac{1}{2} & 0 & \frac{1}{2} & \\
 1 & 0 & 0 & 1 \\
 \hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
 \end{array}. \quad (48)$$

Az előző fejezetben megismert módszerek is Runge-Kutta típusúak voltak, csak az implicit Euler és trapéz módszer implicit Runge-Kutta típusúak, amiknek a Butcher-táblázata nem csak alsó háromszögmátrix, hanem telimátrix.

Az értékeket úgy állítjuk be, hogy a kezdetiérték 1 legyen, a paraméternek pedig $\lambda = 0$ értéket adunk, mert különben ha τ kicsi és λ nagy (lásd 5. ábra), a módszerek nem működnek jól (mi csak explicit Runge-Kutta módszereket vizsgáltunk, implicit módszerekre nem adna ilyen rossz eredményt λ nagy értékeire sem, lásd az 1. táblázatban az implicit Euler és a trapéz-módszer eredményeit). Így egy harmadfokú polinomot becslünk, a lépésköz felezgetésével a 2. táblázatban megfigyelhetjük, hogyan változik a módszerek pontossága.

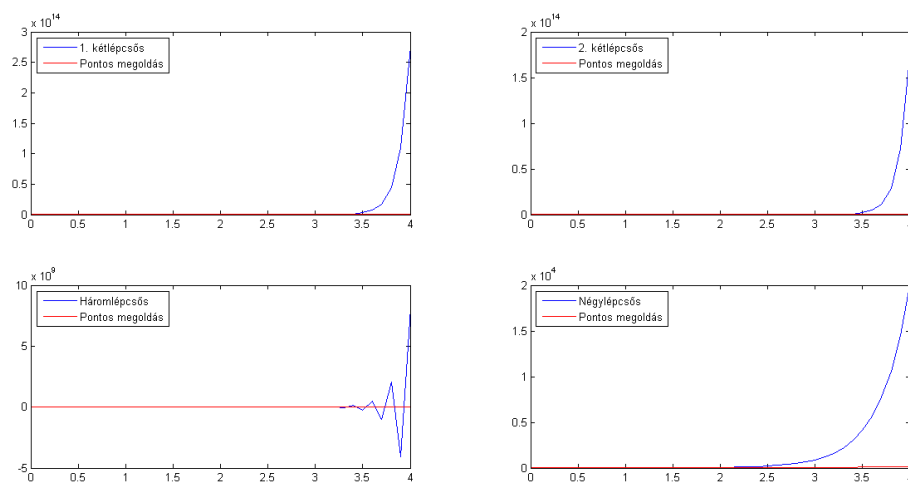
Mind a négy módszer nagyon jó közelítést ad (lásd 6. ábra), a lépésköz felezésével az

2. táblázat. Runge-Kutta módszerek hibája τ felezésével

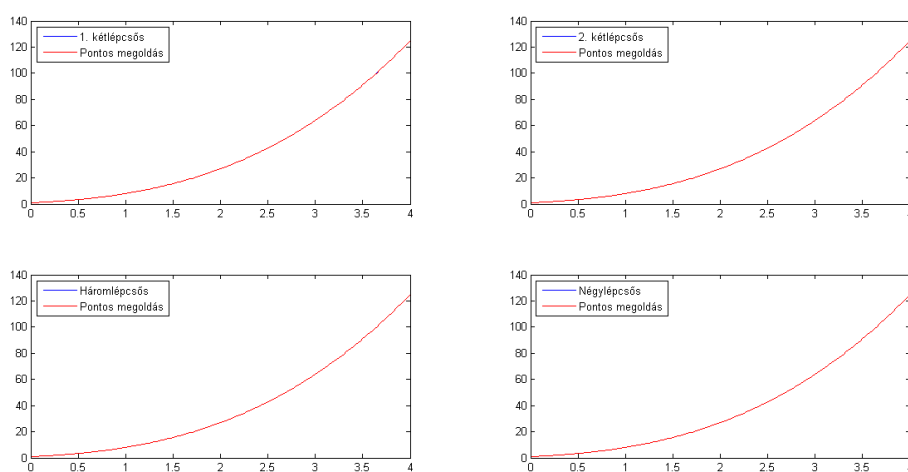
	$\tau = 0,1$	$\tau = 0,05$	$\tau = 0,025$
1. kétlépéses	0,0200	0,0050	0,0012
2. kétlépéses	$2,84 \cdot 10^{-14}$	$1,42 \cdot 10^{-14}$	$< 10^{-16}$
Háromlépéses	$1,42 \cdot 10^{-14}$	$1,14 \cdot 10^{-13}$	$8,53 \cdot 10^{-14}$
Négylépéses	$2,84 \cdot 10^{-14}$	$2,84 \cdot 10^{-14}$	$1,42 \cdot 10^{-14}$

összes becslés pontosabb lesz, de az eredeti célunk nem jött be, azaz annak ellenére, hogy magasabbrendű módszereket tudtunk előállítani, azok csak kevés esetben adtak jó közelítő megoldást.

5. ábra. Runge-Kutta módszerek, $\lambda = 30$, $\tau = 0,1$



6. ábra. Runge-Kutta módszerek, $\lambda = 0$, $\tau = 0,025$



5. Többlépéses módszerek

Az eddig vizsgált módszereknél mindig úgy jártunk el, hogy t_0 -ból indulva próbáltuk meghatározni a t_1 pontbeli értéket, felhasználva azt, hogy ismerjük y_0 kezdeti értéket. Ezután t_1 -ből próbáltunk meg t_2 -beli értékre következtetni, y_1 ismeretében, miközben y_0 értéket nem vettük figyelembe, kidobtuk, és így tovább. Ezeket neveztük egylépéses módszereknek. A többlépéses módszereknél azt az ötletet vesszük figyelembe, hogy az adott pontbeli megoldás, ne csak az eggyel előző, hanem több korábbi értéktől függjön. A többlépéses módszereket fölírhatjuk általános képlettel:

$$a_0 y_n + a_1 y_{n-1} + \dots + a_p y_{n-p} = \tau(b_0 f(t_n, y_n) + b_1 f(t_{n-1}, y_{n-1}) + \dots + b_q f(t_{n-q}, y_{n-q})), \quad (49)$$

ahol az a_i -k és b_i -k konstansok ($a_0 \neq 0$). Ha $b_0 = 0$, a módszer explicit, ha $b_0 \neq 0$, akkor implicit.

5.1. Definíció. *Egy többlépéses módszer első és második karakterisztikus polinomjának nevezzük a*

$$\varrho(z) = a_0 z^s + a_1 z^{s-1} + \dots + a_s$$

$$\sigma(z) = b_0 z^s + a_1 z^{s-1} + \dots + b_s$$

polinomokat, ahol $s := \max\{p, q\}$.

5.2. Definíció. *Azt mondjuk, hogy egy p polinomra teljesül a gyökfeltétel, ha $p(z^*) = 0$ esetén $|z^*| \leq 1$, ahol $|z^*| = 1$ esetén z^* egyszeres gyöke p -nek.*

5.1. Tétel. *Tegyük fel, hogy egy s -lépéses módszernél az $y_1, y_2, \dots, y_{s-1}$ értékek hibája nulla-hoz tart, ha $\tau \rightarrow 0$. Ekkor az s -lépéses módszer pontosan akkor konvergens, ha konzisztens és a $\varrho(z)$ polinomra teljesül a gyökfeltétel.*

A következő módszerekre megvizsgáljuk, hogy teljesül-e a gyökfeltétel, kiszámoljuk a konzisztencia rendjüket, és így megkapjuk az egyes módszerekre a konvergencia rendjét.

5.1. A középpont szabály

A középpont szabálynál a közelítő értéket egy pontban, a kettővel korábbi értékkel együtt határozzuk meg (azaz $s = 2$), a két érték közti pontban számolt derivált segítségével.

$$y_n - y_{n-2} = 2\tau f(t_{n-1}, y_{n-1}) \quad (50)$$

A középpont szabályra teljesül a gyökfeltétel, az első karakterisztikus polinomjának,

$$\varrho(z) = z^2 - 1$$

két gyöke van, 1 és -1 , és ezek mind egyszeres gyökök.

Ha kiszámoljuk a $\psi_n^{(1)}$ értéket, azt fogjuk látni, hogy ez a módszer másodrendben konzisztens.

Az egyszerűség kedvéért a (50) egyenletet átírhatjuk egy ekvivalens formájába:

$$y_{n+1} - y_{n-1} = 2\tau f(t_n, y_n).$$

A $\psi_n^{(1)}$ értéket most már könnyebben kiszámolhatjuk, a képletbe a pontos megoldást helyettesítve, és a megfelelő értékek helyébe az τ Taylor-felbontásos alakjukat írva:

$$\begin{aligned}\psi_n^{(1)} &= -\frac{u(t_{n+1}) - u(t_{n-1}))}{\tau} + 2f(t_n, u(t_n)) \\ \psi_n^{(1)} &= -\frac{1}{\tau} \left(u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2} u''(t_n) + \frac{\tau^3}{6} u'''(t_n) + O(\tau^4) \right. \\ &\quad \left. - u(t_n) + \tau u'(t_n) - \frac{\tau^2}{2} u''(t_n) + \frac{\tau^3}{6} u'''(t_n) + O(\tau^4) \right) + 2u'(t_n) \\ \psi_n^{(1)} &= -2u'(t_n) - \frac{\tau^2}{3} u'''(t_n) + O(\tau^3) + 2u'(t_n) \\ \psi_n^{(1)} &= -\frac{\tau^2}{3} u'''(t_n) + O(\tau^3) = O(\tau^2).\end{aligned}\tag{51}$$

Tehát a középpont szabály konvergencia, a konvergencia rendje 2.

5.2. Az Adams-Bashforth és Adams-Moulton módszerek

Az

$$y_n - y_{n-1} = \tau(b_0 f_n + b_1 f_{n-1} + \dots + b_s f_{n-s})$$

többlépéses módszert Adams-Bashforth⁹ módszernek nevezzük, ha $b_0 = 0$, és ilyenkor a módszer explicit, az implicit módszert, amikor $b_0 \neq 0$, Adams-Moulton¹⁰ módszernek nevezzük. Ahhoz, hogy meghatározzuk a konstansok értékeit, és ezzel együtt magát a módszert, az egyes osztópontok közötti közelítő értéket integrálással fogjuk meghatározni, amit majd Newton¹¹-interpolációval fogunk közelíteni.

Definiáljuk az osztott differencia fogalmát:

$$\begin{aligned}f[t_1] &:= f(t_1), \\ f[t_1, t_2] &:= \frac{f(t_2) - f(t_1)}{t_2 - t_1},\end{aligned}$$

tegyük fel, hogy $f[t_1, t_2, \dots, t_k]$ -t már definiáltuk, ekkor

$$f[t_1, t_2, \dots, t_{k+1}] := \frac{f[t_2, \dots, t_{k+1}] - f[t_1, \dots, t_k]}{t_{k+1} - t_1}.$$

⁹John Couch Adams (1819-1892) és Francis Bashforth (1819-1912), brit matematikusok

¹⁰Forest Ray Moulton (1872-1952), amerikai csillagász

¹¹Sir Isaac Newton (1643-1727), angol matematikus, csillagász

5.2.1. Kétlépéses Adams-Bashforth módszer

Keressünk kétlépéses Adams-Bashforth módszert, ami

$$y_n - y_{n-1} = \tau(b_1 f_{n-1} + b_2 f_{n-2})$$

alakú lesz. Az $y_n - y_{n-1} = y(t_n) - y(t_{n-1})$ kifejezést Newton-interpolációs polinommal közelítjük:

$$y_n - y_{n-1} \approx u_n - u_{n-1} = \int_{t_{n-1}}^{t_n} u'(t) dt = \int_{t_{n-1}}^{t_n} f(t, u(t)) dt = \int_{t_{n-1}}^{t_n} F(t) dt \approx \int_{t_{n-1}}^{t_n} P(t) dt,$$

ahol $P(t)$ lesz az interpolációs polinom. Mivel explicit módszert szeretnénk kapni, ezért csak a t_{n-2}, t_{n-1} pontok között interpolálunk.

$$\begin{aligned} P(t) &= F(t_{n-1}) + F[t_{n-1}, t_{n-2}](t - t_{n-1}) \\ y(t_n) - y(t_{n-1}) &\approx \int_{t_{n-1}}^{t_n} F(t_{n-1}) + F[t_{n-1}, t_{n-2}](t - t_{n-1}) dt \\ &= \tau F(t_{n-1}) + F[t_{n-1}, t_{n-2}] \int_{t_{n-1}}^{t_n} (t - t_{n-1}) dt \end{aligned}$$

Csak az integrált tekintve:

$$\begin{aligned} \int_{t_{n-1}}^{t_n} (t - t_{n-1}) dt &= \left[\frac{t^2}{2} - t \cdot t_{n-1} \right]_{t_{n-1}}^{t_n} = \frac{t_n^2}{2} - t_n \cdot t_{n-1} - \frac{t_{n-1}^2}{2} + t_{n-1}^2 \\ &= \frac{t_n^2}{2} - t_n^2 + \tau t_n + \frac{t_n^2}{2} - \frac{2\tau t_n}{2} + \frac{\tau^2}{2} = \frac{\tau^2}{2}. \end{aligned}$$

Ezt visszahelyettesítve az előbbi képletbe:

$$\begin{aligned} y(t_n) - y(t_{n-1}) &= \tau F(t_{n-1}) + \frac{\tau^2}{2} F[t_{n-1}, t_{n-2}] = \tau f(t_{n-1}) + \frac{\tau^2}{2} \cdot \frac{f_{n-1} - f_{n-2}}{\tau} \\ y(t_n) - y(t_{n-1}) &= \frac{\tau}{2} (3f_{n-1} - f_{n-2}). \end{aligned} \tag{52}$$

A (52) egyenlettel megkaptuk a kétlépéses Adams-Bashforth módszer képletét, ha kiszámoljuk a konzisztencia rendjét, azt kapjuk, hogy a módszer másodrendű.

$$\begin{aligned} \psi_n^{(1)} &= -\frac{u(t_{n+1}) - u(t_n)}{\tau} + \frac{3}{2}u'(t_n) - \frac{1}{2}u'(t_{n-1}) \\ &= -\frac{1}{\tau} \left(u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2}u''(t_n) + \frac{\tau^3}{6}u'''(t_n) - u'(t_n) \right) + \frac{3}{2}u'(t_n) \\ &\quad - \frac{1}{2} \left(u'(t_n) - \tau u''(t_n) + \frac{\tau^2}{2}u'''(t_n) + O(\tau^3) \right) \\ &= -\frac{\tau^2}{12}u'''(t_n) + O(\tau^3) = O(\tau^2) \end{aligned} \tag{53}$$

Erre a módszerre igaz a gyökfeltétel, a konzisztencia rendje 2, emiatt a módszer konvergencia is és a konvergencia rendje 2.

5.2.2. Kétlépéses Adams-Moulton módszer

Most az előző módszer implicit változatát szeretnénk elkészíteni, a

$$y_n - y_{n-1} = \tau(b_0 f_n + b_1 f_{n-1} + b_2 f_{n-2})$$

egyenlet együtthatóinak meghatározásával, ehhez az $y_n - y_{n-1}$ értéket most is Newton-interpolációval közelítjük.

$$\begin{aligned} y_n - y_{n-1} &\approx \int_{t_{n-1}}^{t_n} F(t) dt = F(t_n)(t_n - t_{n-1}) + F[t_n, t_{n-1}](t_n - t_{n-1})^2 \\ &\quad + \frac{1}{2} F[t_n, t_{n-1}, t_{n-2}](t_n - t_{n-1})^3 \\ &= \tau F(t_n) + F[t_n, t_{n-1}] \int_{t_{n-1}}^{t_n} (t - t_{n-1}) dt + \frac{1}{2} F[t_n, t_{n-1}, t_{n-2}] \int_{t_{n-1}}^{t_n} (t - t_{n-1})^2 dt \end{aligned}$$

Az integrálokat kiszámítva,

$$\begin{aligned} \int_{t_{n-1}}^{t_n} (t - t_{n-1}) dt &= \frac{\tau^2}{2} \\ \int_{t_{n-1}}^{t_n} (t - t_{n-1})^2 dt &= \frac{\tau^3}{6} \end{aligned}$$

majd visszahelyettesítve az interpolációs képletbe,

$$\begin{aligned} y_n - y_{n-1} &= \tau f_n - \frac{\tau^2}{2} F[t_n, t_{n-1}] - \frac{\tau^3}{6} F[t_n, t_{n-1}, t_{n-2}] \\ &= \tau f_n - \frac{\tau^2}{2} \cdot \frac{f_n - f_{n-1}}{\tau} - \frac{\tau^3}{6} \cdot \frac{f_n - 2f_{n-1} + f_{n-2}}{\tau^2} \end{aligned}$$

megkapjuk a kétlépéses Adams-Moulton módszer képletét

$$y_n - y_{n-1} = \frac{\tau}{12}(5f_n + 8f_{n-1} - f_{n-2}). \quad (54)$$

Az implicit Adams-Moulton módszer konzisztencia rendjét, az előző módszeréhez hasonlóan számolhatjuk ki:

$$\begin{aligned} \psi_n^{(1)} &= -\frac{u(t_{n+1}) - u(t_n)}{\tau} + \frac{5}{12}u'(t_{n+1}) + \frac{8}{12}u'(t_n) - \frac{1}{12}u'(t_{n-1}) \\ &= -u'(t_n) - \frac{\tau}{2}u''(t_n) - \frac{\tau^2}{6}u'''(t_n) + \frac{5}{12}u'(t_n) + \frac{5\tau}{12}u''(t_n) + \frac{5\tau^2}{24}u'''(t_n) + \frac{8}{12}u'(t_n) \\ &\quad - \frac{1}{12}u'(t_n) + \frac{\tau}{12}u''(t_n) - \frac{\tau^2}{24}u'''(t_n) + O(\tau^3) \\ \psi_n^{(1)} &= O(\tau^3). \end{aligned} \quad (55)$$

Mivel erre a módszerre is igaz a gyökfeltétel, ezért a kétlépéses Adams-Moulton módszer konvergencia rendje 3.

5.3. A Nyström és a Milne-Simpson módszerek

A két módszert általánosan leíró képlet a következő:

$$y_n - y_{n-2} = \tau(b_0 f_n + b_1 f_{n-1} + \dots + b_s f_{n-s}), \quad (56)$$

ha $b_0 = 0$, akkor Nyström¹² módszernek hívjuk, ha $b_0 \neq 0$, akkor Milne-Simpson¹³ módszernek. Az előző esetben explicit, az utóbbiban implicit a módszer. Egy konkrét esetet vizsgálva, például ha az (56) egyenletet $s = 2$ esetben nézzük, akkor kapunk egy implicit Milne-Simpson módszert, aminek az együtthatóit kiszámolva kapjuk a Simpson-formulának elnevezett módszert:

$$y_n - y_{n-2} = \frac{2\tau}{6}(f_n + 4f_{n-1} + f_{n-2}).$$

A Nyström módszer egyik speciális esete a 5.1. fejezetben vizsgált középpont szabály.

5.4. A prediktor-korrektor módszerek

Ebben a fejezetben eddig többlépéses implicit és explicit módszereket is vizsgáltunk, most viszont egy olyan módszerrel foglalkozunk, ami mindkét módszert felhasználja. Ugyanis például az Adams-Bashforth módszereket nem szokták magukban alkalmazni, mindig egy implicit módszerrel együtt használják. A prediktor-korrektor módszer lényege az, hogy amikor a t_n pontból becsülünk a t_{n+1} pontbeli értékre, akkor először egy explicit módszerrel kapunk egy értéket y_{n+1} -re, ezt jelöljük y_{n+1}^0 -val, majd ezt beírva f_{n+1} -be, egy implicit módszert használva javítunk a t_{n+1} -beli becslésen. Ha például a már korábban a 5.2.1. és 5.2.2. pontokban vizsgált explicit Adams-Bashforth és implicit Adams-Moulton módszereket használjuk, az alábbiak szerint kell eljárunk:

$$\begin{array}{ll} \mathbf{P} & y_{n+1}^0 = y_n + \frac{\tau}{2}(3f(t_n, y_n) - f(t_{n-1}, y_{n-1})) \\ \mathbf{E} & f_{n+1}^0 = f(t_{n+1}, y_{n+1}^0) \\ \mathbf{C} & y_{n+1}^1 = y_n + \frac{\tau}{12}(5f_{n+1}^0 + 8f_n - f_{n-1}) \\ \mathbf{E} & f_{n+1}^1 = f(t_{n+1}, y_{n+1}^1), \end{array}$$

P a prediktor módszert jelöli, **C** a korrektor (corrector) módszert, **E** az értékadást (evaluation). A jobb közelítés érdekében a második és harmadik lépést lehet iterálni, ekkor **P(EC)^mE** módszert kapunk, ahol m az iterációk számát jelöli.

A prediktor-korrektor módszer rendjére az alábbi tétel igaz:

5.2. Tétel. *Legyen egy prediktor-korrektor módszer prediktor tagjának rendje p^* , a korrektor tagé p . Ekkor az egész módszer rendjére, q_m -re, igaz, hogy*

$$q_m = \min(p, p^* + m),$$

¹²Evert J. Nyström (1895-1960), finn matematikus

¹³Thomas Simpson (1710-1761), brit matematikus; William Edmund Milne (1890-1971), amerikai matematikus

ahol m az az iterációk számát jelöli.

Ebből következik az, hogy az iterációs számot általában egynek szokták választani, mert legtöbbször olyan prediktor-korrektor párt használnak, amelyek rendje megegyezik, vagy maximum eggyel tér el. Az előbb felírt módszer rendje 3, hiszen a kétlépéses Adams-Bashforth rendje 2, az Adams-Moulton rendje 3.

5.5. Többlépéses módszerek tesztelése

A többlépéses módszerek közül, a szakdolgozatban is bővebben kifejtett, középpont szabályt, kétlépéses Adams-Bashforth és Adams-Moulton és a kétlépéses Adams-Bashforth és szintén kétlépéses Adams-Moulton módszerekből álló prediktor-korrektor módszert fogjuk tesztelni. Az első kettő rendje kettő, az Adams-Moultonnak és a prediktor-korrektornak viszont három, ezért ezektől a módszerektől jobb eredményeket várunk. A programot a szokásos paraméterek kombinációival futtatjuk, $y_0 = 10$, $\tau = 0,1$ vagy $0,05$ vagy $0,025$ és $\lambda = 0, 15, 30$.

3. táblázat. Többlépéses módszerek hibája különböző τ -val és λ -val

		Középpont szabály	Adams- Bashforth	Adams- Moulton	Prediktor- korrektor
$\tau = 0,1$	$\lambda = 0$	0,0400	0,0975	$< 10^{-16}$	$2,84 \cdot 10^{-14}$
	$\lambda = 15$	$1,13 \cdot 10^{20}$	$1,30 \cdot 10^9$	$1,42 \cdot 10^{-14}$	0,0051
	$\lambda = 30$	$6,22 \cdot 10^{30}$	$2,96 \cdot 10^{23}$	$8,61 \cdot 10^{-9}$	$5,35 \cdot 10^{29}$
$\tau = 0,05$	$\lambda = 0$	0,0100	0,0247	$2,84 \cdot 10^{-14}$	$< 10^{-16}$
	$\lambda = 15$	$1,20 \cdot 10^{23}$	$4,17 \cdot 10^{-4}$	$2,84 \cdot 10^{-14}$	$5,18 \cdot 10^{-5}$
	$\lambda = 30$	$6,44 \cdot 10^{40}$	$1,82 \cdot 10^{18}$	$1,42 \cdot 10^{-14}$	$7,64 \cdot 10^{-5}$
$\tau = 0,025$	$\lambda = 0$	$2,50 \cdot 10^{-3}$	$6,20 \cdot 10^{-3}$	$5,68 \cdot 10^{-14}$	$2,48 \cdot 10^{-14}$
	$\lambda = 15$	$7,31 \cdot 10^{23}$	$1,04 \cdot 10^{-4}$	$2,48 \cdot 10^{-14}$	$2,93 \cdot 10^{-6}$
	$\lambda = 30$	$1,45 \cdot 10^{47}$	$5,21 \cdot 10^{-5}$	$2,48 \cdot 10^{-14}$	$6,48 \cdot 10^{-6}$

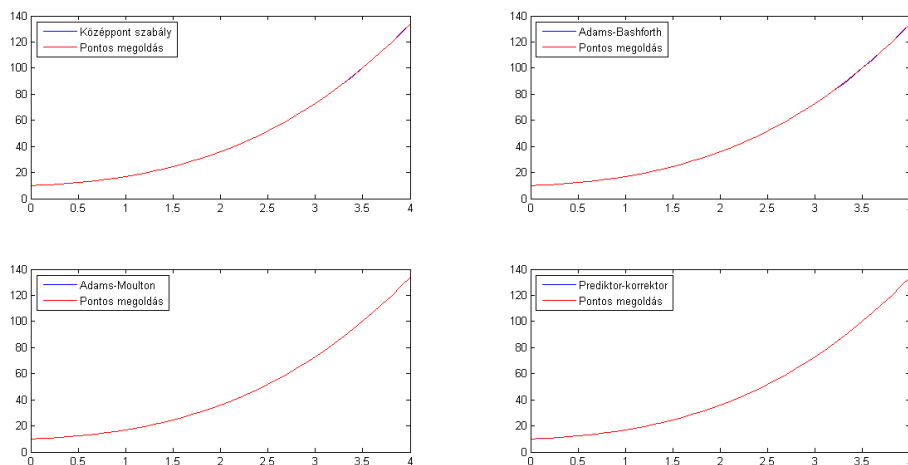
A középpont szabály szinte az összes beállításra rossz közelítést ad (lásd 3. táblázat), egyedül $\lambda = 0$ -ra ad jó közelítést (lásd 7. ábra). A többi módszer nagyon jól közelíti a pontos megoldást ha τ és λ szorzata kicsi, a másodfokú Adams-Bashforthnál mindig jobb a két harmadfokú. Az egy lépéses vagy a Runge-Kutta módszereknél tapasztaltuk, hogy bizonyos paraméterekkel, amikor τ és λ szorzata nagy, az explicit módszerek nagyon rossz közelítést adnak és az implicitek jobbak lesznek. Ez a többlépéses módszereknél is megfigyelhető, de ezek a módszerek még érzékenyebbek a paraméterek megválasztására, például ezt tapasztaljuk már $\tau = 0,1$ és $\lambda = 15$ -nél (lásd 8. ábra), vagy $\tau = 0,05$ és $\lambda = 30$ -

nál is (lásd 10. ábra), sőt $\lambda = 30$, $\tau = 0,1$ -nél a prediktor-korrektor módszer is elromlik (lásd 9. ábra).

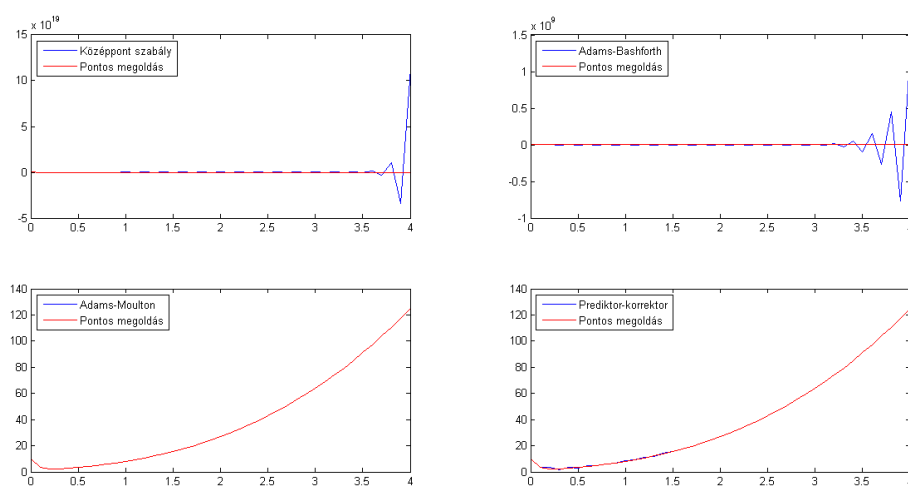
Az implicit Adams-Moulton módszer kicsit ellenállóbb, mert hasonlóan jártunk el, mint az implicit Euler módszernél, azaz az egyenlet linearitását kihasználva ki tudtuk fejezni y_{n+1} -et. Ez viszont általában nem lehetséges, ezért használjuk inkább a prediktor-korrektor módszereket. Azért az Adams-Moulton módszer sem lesz minden esetben jó, λ növelésével ez a módszer is elkezd romlani. Ha $\tau = 0,1$ mellett a λ -t 65-nek választjuk, a közelített megoldás elkezd kicsit oszcillálni a pontos megoldás körül (lásd 11. ábra), hibája a 4-ben 4,9761 lesz, $\lambda = 100$ -ra pedig teljesen elromlik (lásd 12. ábra), itt a hibája $3,2 \cdot 10^3$ lesz. Ezekkel a beállításokkal már az összes általunk vizsgált módszer elszáll, legyen az egy- vagy többlépéses, két módszer viszont még mindig jól közelíti a pontos megoldást¹⁴. Ez a két módszer a két implicit elsőrendű, az implicit Euler és a trapéz-szabály, amelyeknél a hiba $\lambda = 65$ esetén 0,0228 illetve $7,6 \cdot 10^{-5}$, $\lambda = 100$ esetén még mindig kevés, 0,0149 és $5 \cdot 10^{-5}$.

¹⁴Ez igaz lesz több általunk nem vizsgált módszerre is, például az implicit Runge-Kuttákra vagy a Gear-módszerekre.

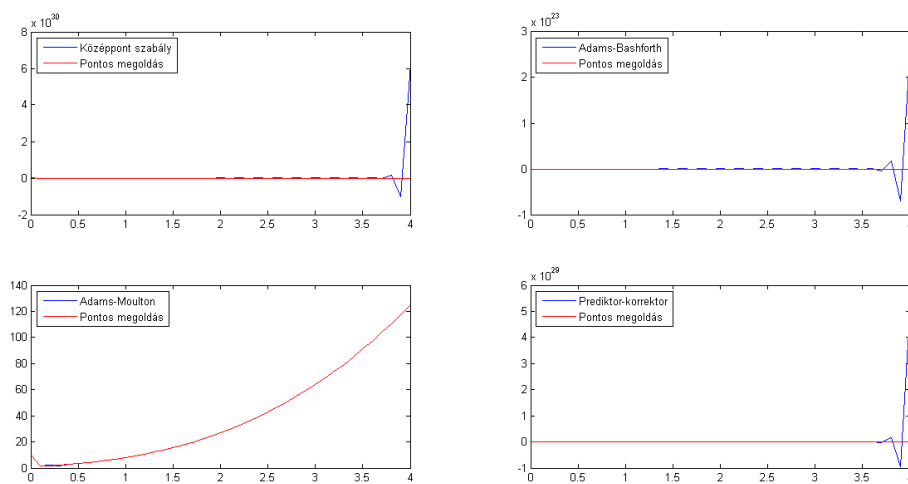
7. ábra. Többlépéses módszerek, $\lambda = 0$, $\tau = 0,1$



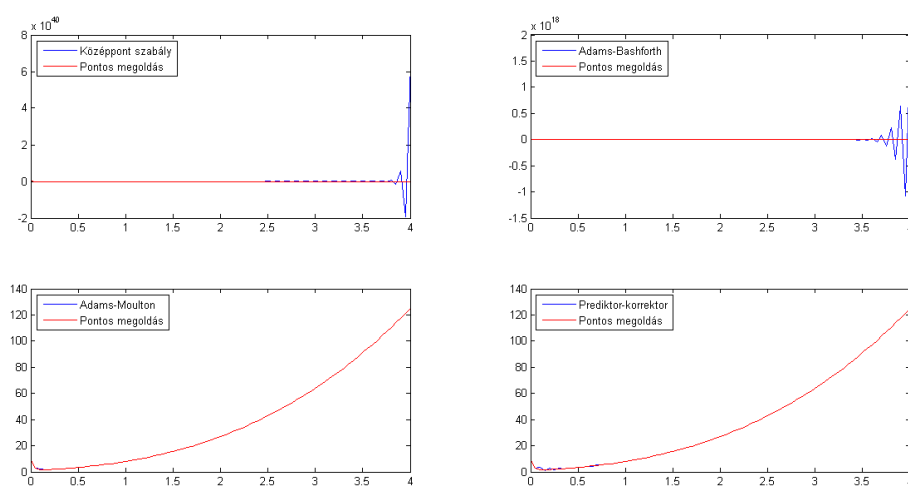
8. ábra. Többlépéses módszerek, $\lambda = 15$, $\tau = 0,1$



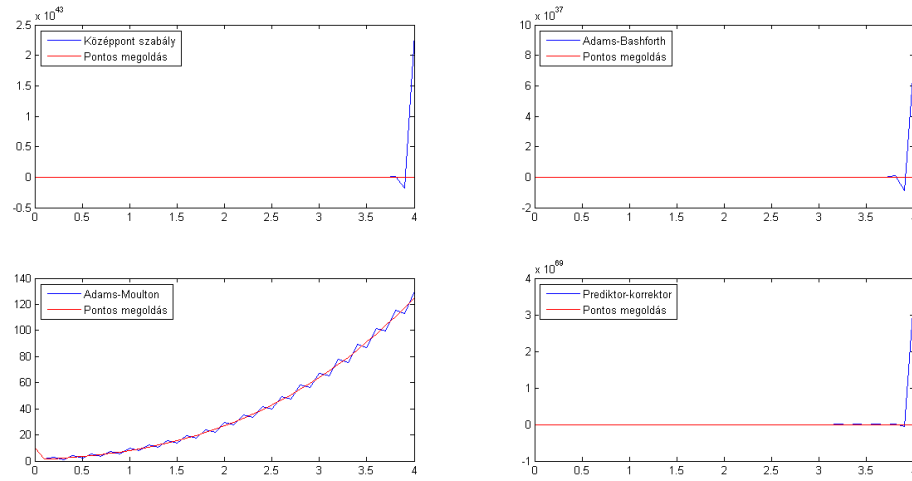
9. ábra. Többlépéses módszerek, $\lambda = 30$, $\tau = 0,1$



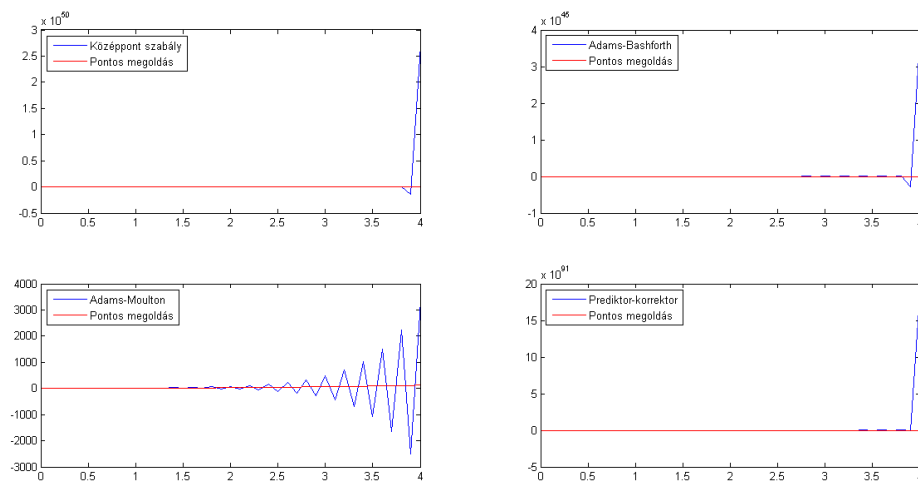
10. ábra. Többlépéses módszerek, $\lambda = 30$, $\tau = 0,05$



11. ábra. Többlépéses módszerek, $\lambda = 65$, $\tau = 0,1$



12. ábra. Többlépéses módszerek, $\lambda = 100$, $\tau = 0,1$



6. Összefoglalás

A célunk az volt, hogy egy

$$u'(t) = f(t, u(t))$$

alakú, u_0 kezdeti értékű közönséges differenciálegyenlet megoldására keressünk numerikus módszereket, amik a pontos megoldást minél jobban közelítik. A szakdolgozat során megismerkedtünk egylépéses módszerekkel, Runge-Kutta és többlépéses módszerekkel is, megvizsgáltuk a módszerek rendjét és egy konkrét példán vizsgáltuk a módszerek működését. Megfigyeltük az összefüggést a rend és a módszer pontossága között, valamint a paraméterek, ugymint a lépésszám és a λ érték, hatását a numerikus módszerek működésére.

A tesztfeladatunk speciális volt, a pontosság megfigyelésére igazából csak a $\lambda = 0$ választás volt alkalmas, nagy λ -ra merev lett az egyenlet, ezért ebben az esetben az implicit módszerek rendtől függetlenül jobb eredményt adtak az explicitéknél, sőt az implicit Euler módszer jobb lett mint egy általános többlépéses implicit módszer.

A merev differenciálegyenletekre alkalmas módszerek vizsgálata meghaladja a dolgozat kereteit, ilyen módszerek a már korábban említett implicit Runge-Kutták és a Gear-módszerek. A merev rendszerek olyan egyenletek, ahol a lépésköz választását sokkal inkább a stabilitás határozza meg, mint a pontossági követelmények.

Hivatkozások

- [1] F. R. Giordano, M. D. Weir, W. P. Fox: *A first course in mathematical modeling*, Brooks/Cole.
- [2] Stoyan Gisbert: *Numerikus matematika mérnököknek és programozóknak*, TypoTeX Kft.
- [3] Stoyan Gisbert - Takó Galina: *Numerikus módszerek II.*, ELTE - TypoTeX kiadó.
- [4] Peter Henrici: *Numerikus analízis*, Műszaki Könyvkiadó.
- [5] Arie Iserles: *A first course in the numerical analysis of differential equations*, Cambridge University Press.