Eötvös Loránd Tudományegyetem
Természettudományi Kar

Közönséges differenciálegyenletek numerikus megoldási módszerei

BSc Szakdolgozat

Készítette:

Takács Krisztina

Matematika BSc, Matematikai elemző szakirány

Témavezető: Havasi Ágnes
Eötvös Loránd Tudományegyetem, Természettudományi Kar
Alkalmazott Analízis és Számításmatematikai Tanszék

Budapest, 2018

Köszönetnyilvánítás

Ezúton is szeretnék köszönetet mondani témavezetőmnek, Havasi Ágnesnek a sok segítségért és hasznos tanácsért, amit a dolgozat megírása közben kaptam tőle.

Hálával tartozom Faragó István tanár úrnak, aki érdekes tanóráival megkedveltette velem a témakört.

Külön szeretnék köszönetet mondani szüleimnek és testvéremnek, akik mindig hittek bennem és támogattak tanulmányaim során.

Tartalomjegyzék

1. Bevezetés **4**

2. A közönséges differenciálegyenlet fogalma **6**

3. Véges differenciás módszerek **9**

3.1. Az explicit Euler-módszer 10

3.2. Az explicit Euler-módszer vizsgálata 11

3.3. Az explicit Euler-módszer konvergenciája 13

3.4. Az implicit Euler-módszer 15

3.5. Az implicit Euler-módszer vizsgálata 15

3.6. A trapézmódszer vizsgálata 16

3.7. A θ-módszer 17

3.8. Példa 17

4. A numerikus módszerek stabilitása **20**

4.1. Az explicit Euler-módszer stabilitási tartománya 21

4.2. Az implicit Euler-módszer stabilitási tartománya 22

5. A Richardson-extrapoláció **24**

5.1. Bevezetés 24

5.2. A Richardson-extrapoláció pontossági rendje 26

5.3. A pontosság további növelése 27

5.4. Hibabecslés és a lépésköz automatikus szabályozása 28

5.5. Az aktív és a passzív Richardson-extrapoláció 30

6. Összefoglalás **32**

7. Függelék **33**

1. Bevezetés

A differenciálegyenlet, mint fogalom a sebesség fogalmából született. A differenciálegyenletek segítségével különböző természettudományi, műszaki vagy közgazdasági folyamatokat tudunk modellezni. Attól függően, hogy az ismeretlen függvény egy- vagy többváltozós, beszélhetünk közönséges, illetve parciális differenciálegyenletekről. Egy jó modelltől mindig elvárjuk, hogy korrekt kitűzésű legyen, azaz teljesüljön rá az egzisztencia (azaz létezzen megoldása), az unicitás (azaz csak egyetlen megoldása létezzen), és a stabilitás (a megoldás folytonosan függjön a feladatot leíró paraméterektől) feltétele.

Az egyszerűbb differenciálegyenleteket analitikusan is meg tudjuk oldani, de az esetek többségében ezt a pontos megoldást nem tudjuk előállítani, így numerikus módszerekkel, közelítőleg oldjuk meg a feladatokat. A dolgozatban a közönséges differenciálegyenletek numerikus megoldási módszereivel foglalkozunk.

A második fejezetben a differenciálegyenletekkel kapcsolatos fontosabb fogalmakat és tételeket ismertetjük, definiáljuk a közönséges differenciálegyenletek kezdetiérték-feladatát. A harmadik fejezetben a véges differenciás numerikus módszerekkel, azok levezetésével foglalkozunk. Definiáljuk a rácsháló fogalmát, bevezetjük a különböző egylépéses módszereket, majd a konzisztencia fogalmát, és meghatározzuk ezen módszerek konzisztenciarendjét. Végül egy példán keresztül összehasonlítjuk az implicit és az explicit Euler-módszer alkalmazásával kapott megoldásokat a pontos megoldással. A negyedik fejezetben a stabilitásról lesz szó. Az alapvető fogalmak bevezetése után meghatározzuk két fontosabb numerikus módszer stabilitási tartományát.

Az ötödik fejezetben a Richardson-extrapolációval foglalkozunk. Ezt a számítási eszközt Lewis Fry Richardson angol matematikusról nevezték el, aki többek között a légköri folyamatok modellezésével is foglalkozott. Az első előrejelzési próbálkozása nem volt sikeres, de munkásságával megalapozta a mai időjárás-előrejelzést [10, 11]. A Richardson-extrapoláció egy rendkívül népszerű eszköz a differenciálegyenlet(rendszer)ek numerikus megoldása során. Használatával egy p-ed rendben pontos módszerből $(p+1)$-ed rendben pontos módszert nyerhetünk. A numerikus módszerek esetében sokszor abba a problémába ütközünk, hogy a megoldás, amelyet egy bizonyos lépésközzel kaptunk, nem kellőképpen pontos, így a megoldást elölről kell kezdenünk egy másik, finomabb rácsháló. A koráb-

bi számításaink így kárba vesznek. Richardson azonban kétféle lépésközzel alkalmazta ugyanazt a numerikus módszert, majd ezek megfelelő lineáris kombinációját vette, így pontosabb eredményt kapott. Ebben a fejezetben levezetjük a Richardson-extrapoláció pontossági rendjét, majd ismételten alkalmazzuk az extrapolációt, így egy új módszert kapunk, amely még pontosabb approximációt biztosít. Ezután becslést adunk a hibára, végül röviden ismertetjük a Richardson-extrapoláció két változatát: a passzív és az aktív Richardson-extrapolációt.

2. A közönséges differenciálegyenlet fogalma

A közönséges differenciálegyenlet olyan egyenlet, amely egy ismeretlen, egyváltozós függvény és annak deriváltja közötti kapcsolatot írja le. Ebben a fejezetben összegyűjtjük a közönséges differenciálegyenletek megoldásával kapcsolatos legfontosabb fogalmakat és tételeket.

2.1. Definíció. Legyen $n \in \mathbb{N}$ és $F : \mathbb{R}^{n+2} \to \mathbb{R}$ folytonos függvény. Ekkor n-edrendű közönséges differenciálegyenletnek nevezzük az $F(t, x(t), x'(t), \ldots, x^{(n)}(t)) = 0$ alakú egyenletet.

2.2. Definíció. Az ismeretlen függvény deriváltjai közül az előforduló legmagasabb rendű derivált rendjét a differenciálegyenlet rendjének nevezzük.

2.3. Definíció. Legyen $n \in \mathbb{N}$ és $f : \mathbb{R}^{n+1} \to \mathbb{R}^n$ folytonos függvény. Ekkor elsőrendű közönséges differenciálegyenlet-rendszernek nevezzük az $x'(t) = f(t, x(t))$ alakú differenciálegyenlet-rendszert.

Egy differenciálegyenletet rendszerint nem önmagában oldunk meg, hanem olyan megoldását keressük, amely eleget tesz egy bizonyos kiegészítő feltételnek vagy feltételeknek.

2.4. Definíció. Legyen $G \subset \mathbb{R}^{d+1}$ egy tartomány (azaz G egy összefüggő, nyílt halmaz), $(t_0, u_0) \in G$ egy adott pont ($t_0 \in \mathbb{R}, u_0 \in \mathbb{R}^d$), $f : G \to \mathbb{R}^d$ egy folytonos leképezés. Az

$$u'(t) = f(t, u(t)) \tag{1}$$

$$u(t_0) = u_0 \tag{2}$$

feladatot kezdetiérték-feladatnak, vagy más szóval Cauchy-feladatnak nevezzük.

2.5. Definíció. Az olyan $u : I \to \mathbb{R}^d$ (I egy nyílt intervallum) folytonosan differenciálható függvényt, amelyre

- $\{(t, u(t)) : t \in I\} \subset G$;
- $u'(t) = f(t, u(t)),\ \forall t \in I$;
- $t_0 \in I$ és $u(t_0) = u_0$

az (1), (2) Cauchy-feladat megoldásának nevezzük.

Egy Cauchy-feladat megoldásakor meghatározzuk az összes olyan $u : \mathbb{R} \to \mathbb{R}^d$ függvényt, amely valamely $I \subset \mathbb{R}$ intervallum pontjaiban behelyettesíthető az (1), (2) feladatba, valamint ki is elégíti azt.

Amennyiben a Cauchy-feladat egy természettudományos, műszaki vagy közgazdasági folyamat matematikai modellje, alapvető elvárás, hogy a feladatnak létezzen egyértelmű megoldása. Bizonyos feltételek mellett ez a követelmény biztosítható. Vezessük be az alábbi jelölést:

$$H_{\alpha,\beta}(t_0, u_0) = \{(t, u) : |t - t_0| \leq \alpha, \|u - u_0\|_\infty \leq \beta\} \subset G,$$

ahol α, β pozitív számok. (Tehát $H_{\alpha,\beta}(t_0, u_0)$ egy (t_0, u_0) középű, zárt, $(d+1)$-dimenziós téglalap.) Mivel f folytonos a zárt $H_{\alpha,\beta}(t_0, u_0)$-on, ezért létezik $f(t, u)$-nak maximuma ezen a halmazon. Így értelmes az alábbi valós szám bevezetése: $M = \max_{H_{\alpha,\beta}(t_0,u_0)} \|f(t, u)\|_\infty$. Ekkor $\forall\ t, |t - t_0| \leq \min\{\alpha, \frac{\beta}{M}\}$ esetén létezik az (1), (2) feladatnak megoldása.

2.1. Tétel. *Ha az f függvény a második változójában Lipschitz-tulajdonságú a $H_{\alpha,\beta}(t_0, u_0)$ halmazon, azaz létezik olyan $L \geq 0$ állandó, amelyre $\forall\ (t, u_1), (t, u_2) \in H_{\alpha,\beta}(t_0, u_0)$ pontban $\|f(t, u_1) - f(t, u_2)\|_\infty \leq L\|u_1 - u_2\|_\infty$ teljesül, akkor az* (1), (2) *Cauchy-feladatnak létezik egyértelmű megoldása.*

A továbbiakban mindig feltesszük, hogy a fenti feltételek teljesülnek, azaz létezik olyan $H_{\alpha,\beta}(t_0, u_0) \subset G$ részhalmaz, amelyen f folytonos és a második változójában lipschitzes, azaz létezik egyértelmű megoldása. Legyen $I_0 := \{t \in I : \|t - t_0\| \leq T\}$ intervallum, ahol $T = \min\{\alpha, \frac{\beta}{M}\}$ és a közelítő megoldást ezen az I_0 intervallumon állítjuk elő.

Mivel az alkalmazásokban felmerülő differenciálegyenletek többségében a t változó időt jelent, ezért egy Cauchy-feladat megoldása a rendszer időbeli változását adja meg. Ismerve a rendszer állapotát egy adott t_0 kezdeti időpontban, annak ezt követő, t-beli állapotát szeretnénk meghatározni. Tehát ha $u(t_0)$ ismert, akkor az $u(t)$ függvényt szeretnénk megadni, ahol $t > t_0$. A $t = t_0$ időpontot kezdőpontnak, a megoldásfüggvény t_0-beli értékét kezdeti értéknek, valamint a $(t_0, u(t_0))$ párt kezdeti feltételnek nevezzük.

Legyen a kezdőpont $t_0 = 0$, így a feladat megoldásának értelmezési tartománya a

$[0,T] \subset I$ intervallum, és ekkor a feladat a következő:

$$u'(t) = f(t, u(t)),\ t \in [0,T]$$
$$u(0) = u_0.$$

Célunk az $u(t)$ megoldásfüggvény meghatározása.

A könnyebb áttekinthetőség kedvéért mostantól a $d = 1$ skaláris esetet vizsgáljuk. Ekkor legyen $Q_T := [0,T] \times \mathbb{R} \subset \mathbb{R}^2,\ f : Q_T \to \mathbb{R}$. A továbbiakban az

$$u' = f(t,u)$$
$$u(0) = u_0,\ u_0 \in \mathbb{R}$$

feladatot nevezzük Cauchy-feladatnak, és feltesszük, hogy $f \in C(Q_T)$ és a második változójában lipschitzes függvény, azaz

$$|f(t,u_1) - f(t,u_2)| \leq L|u_1 - u_2|,\ \forall\ (t,u_1),(t,u_2) \in Q_T.$$

Tehát a feladat a következő: keressük azt a megfelelően sima $u : [0,T] \to \mathbb{R}$ függvényt, amelyre

$$u'(t) = f(t, u(t)), \forall\ t \in [0,T] \tag{3}$$
$$u(0) = u_0. \tag{4}$$

Ezt a fejezetet az [1], [2], [4] jegyzetek segítségével dolgoztuk fel.

3. Véges differenciás módszerek

Az előző fejezetben a megoldás létezésére és egyértelműségére vonatkozó feltételeket vizsgáltunk, azonban ezek nem adnak választ annak előállítására. Egy közönséges differenciálegyenlet kezdetiérték-feladatának megoldását általános esetben nem tudjuk meghatározni zárt alakban, ezért a pontos megoldást valamilyen numerikus módszer segítségével közelítő alakban keressük. Ez azt jelenti, hogy értelmezési tartományának egyes pontjaiban a megoldásfüggvény értékeit véges számú lépéssel közelítőleg határozzuk meg. Ebben a fejezetben olyan eljárásokkal foglalkozunk, amelyek valamely rögzített időpontbeli közelítést egy, azt közvetlenül megelőző időpontbeli közelítés felhasználásával határoznak meg. Az ilyen módszereket egylépéses módszereknek nevezzük. Tehát célunk az

$$u' = f(t, u), \ t \in [0, T] \tag{5}$$

$$u(0) = u_0 \tag{6}$$

feladat egylépéses módszerekkel történő közelítő megoldása, ahol $T > 0$ olyan szám, amely mellett az (5), (6) feladatnak létezik egyértelmű, elég sima megoldása a $[0, T]$ intervallumon.

Vezessük be az $\omega_\tau := \{t_n = n \cdot \tau, n = 0, 1, 2, \ldots\}$ ekvidisztáns felosztású rácshálót, ahol $\tau > 0$ adott paraméter. Tehát ω_τ olyan rácsháló, ahol a pontok egyenlő távolságra helyezkednek el egymástól. Célunk egy olyan $y_\tau : \omega_\tau \to \mathbb{R}$ rácsfüggvény meghatározása, amely az ω_τ pontjaiban jól közelíti a (3),(4) Cauchy-feladat pontos $u(t)$ megoldását, azaz $y_\tau(t_n) \approx u(t_n)$, ahol $t_n \in \omega_\tau$. Ehhez bevezetjük az $y_\tau(t_n) =: y_n$, valamint az $u(t_n) =: u_n$ jelöléseket.

Az alábbi kérdések merülhetnek fel:

- Mikor jó a numerikus módszer?
  - ha hatékony, azaz a megoldás előállítása nem igényel túlságosan nagy számítási időt,
  - ha egyszerre konzisztens és stabil, mert a kettő együtt biztosítja a konvergenciát. (Megjegyzés: Ha a numerikus módszer p-ed rendben konzisztens és stabil, akkor p-ed rendben konvergens is. Ez a tétel Lax Péter magyar származású amerikai matematikus nevéhez fűződik.)

- Mit értünk azon, hogy y_n „jól közelíti” u_n-t? ($y_n \approx u_n$)

  Tekintsük az előbb definiált ω_τ rácshálók sorozatát $\tau \to 0$ mellett, és vezessük be a következő definíciókat!

  3.1. Definíció. Azt mondjuk, hogy egy numerikus módszer konvergens a $t^* \in \omega_\tau$ pontban, ha az általa meghatározott y_τ rácsfüggvényre érvényes a

  $$\lim_{\tau \to 0} |y_n - u(t^*)| = 0 \tag{7}$$

  reláció, ahol $n \cdot \tau = t^*$. Ha (7) érvényes a $[0, T]$ intervallum minden t^* pontjában, akkor a numerikus módszert konvergensnek nevezzük ezen az intervallumon.

  3.2. Definíció. Azt mondjuk, hogy a numerikus módszer p-ed rendben konvergens, ha

  $$|y_n - u(t^*)| = \mathcal{O}(\tau^p), \ p \geq 1. \tag{8}$$

  Célunk, hogy minél magasabb p-re tudjuk garantálni ezt a konvergenciát. Minden véges különbséges módszerrel szemben alapvető elvárásunk, hogy a numerikus megoldás a fent definiált értelemben tartson a feladat pontos megoldásához.

- Hogyan határozzuk meg az y_τ rácsfüggvényt?

  Erre a kérdésre a következő alfejezetekben fogunk választ adni, és több módszert is tekintünk a rácsfüggvény meghatározására.

3.1. Az explicit Euler-módszer

Mivel u pontos megoldás, ezért speciálisan $t = t_n$-ben is kielégíti az eredeti (5) feladatot, vagyis

$$u'(t_n) = f(t_n, u(t_n)). \tag{9}$$

Egy függvény deriváltját a differenciahányados-függvény határértékeként értelmeztük, ezért célszerű a deriváltat a differenciahányados értékeivel approximálni. A t_n-beli deriváltat az

$$u'(t_n) \approx \frac{u(t_{n+1}) - u(t_n)}{\tau} \tag{10}$$

jobb oldali véges differenciahányadossal közelíthetjük. Elsőként a megoldás t_1-beli közelítő értékét szeretnénk megkapni. A t_0-beli deriváltra alkalmazva a (9), (10) képletet az

$$f(t_0, u(t_0)) \approx \frac{u(t_1) - u(t_0)}{\tau} \tag{11}$$

összefüggést kapjuk. Mivel célunk az, hogy y_n „jól közelítse" $u(t_n)$-t, azaz $y_n \approx u(t_n)$ teljesüljön, valamint $y_0 = u(t_0)$ adva van a kezdeti feltételből, ezért (11)-et átrendezve

$$u(t_1) \approx f(t_0, u(t_0))\tau + u(t_0) = f(t_0, y_0)\tau + y_0 =: y_1$$

lesz a megoldás t_1-beli értékének közelítése.

Hasonlóan kaphatjuk meg a t_2-beli megoldás közelítését:

$$u(t_2) \approx f(t_1, u(t_1))\tau + u(t_1) \approx f(t_1, y_1)\tau + y_1 =: y_2$$

és így tovább. Így a következő algoritmushoz jutunk:

$$y_{n+1} = y_n + \tau f(t_n, y_n), \ n = 0, 1, 2, \ldots, N-1. \tag{12}$$

Vagyis y_n-ből közvetlenül ki tudjuk számolni y_{n+1}-et. Tehát felírhatunk egy egylépéses iterációt a következő módon:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n), \ n = 0, 1, 2, \ldots, N-1 \tag{13}$$

$$y_0 = u_0. \tag{14}$$

A (13), (14) módszert explicit Euler-módszernek nevezzük.

3.2. Az explicit Euler-módszer vizsgálata

Vezessük be a következő jelölést:

$e_n := y_n - u(t^*)$, azaz $e_n = y_n - u_n$, ahol $u_n = u(t_n) \equiv u(t^*)$.

3.3. Definíció. Az $e_n = y_n - u(t_n)$ rácsfüggvényt hibafüggvénynek vagy globális hibafüggvénynek nevezzük.

A globális hibafüggvény azt mutatja meg, hogy a pontos és a közelítő megoldás milyen messze van egymástól. Helyettesítsük be a (13) rekurzióba az $y_n = e_n + u_n$ képletet, ekkor az alábbi összefüggést kapjuk:

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

Ezt átrendezzük:

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

A jobb oldalhoz először adjuk hozzá az $f(t_n, u_n)$ értéket, majd vonjuk is ki belőle, ekkor a következő egyenlőséget kapjuk:

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

Vezessük be a következő jelöléseket:

$$\Psi_n^{(1)} := -\frac{u_{n+1}-u_n}{\tau} + f(t_n, u_n) \tag{15}$$

$$\Psi_n^{(2)} := -f(t_n, u_n) + f(t_n, e_n + u_n). \tag{16}$$

Az explicit Euler-módszer hibaegyenlete:

$$\frac{e_{n+1}-e_n}{\tau} = \Psi_n^{(1)} + \Psi_n^{(2)},$$

ahol a $\Psi_n^{(1)}$ tagot lokális approximációs hibának nevezzük. Ez azt mutatja meg, hogy a pontos megoldás milyen pontosan elégíti ki a numerikus megoldást meghatározó egyenletet. $\Psi_n^{(1)}$ ugyanis nem más, mint a pontos megoldás behelyettesítése a nullára rendezett (10) képletbe:

$$-\frac{y_{n+1}-y_n}{\tau} + f(t_n, y_n) = 0.$$

3.1. Megjegyzés. Mivel a numerikus megoldás általában különbözik a pontos megoldástól, ezért a $\Psi_n^{(1)}$ lokális approximációs hiba nem egyenlő nullával, tehát ha $e_n \neq 0$, akkor $y_n \neq u_n$.

3.4. Definíció. Azt mondjuk, hogy egy numerikus módszer konzisztens, ha

$$\lim_{\tau \to 0} \Psi_n^{(1)} = 0.$$

Ha $\Psi_n^{(1)} = \mathcal{O}(\tau^p)$, akkor a numerikus módszert p-ed rendben konzisztensnek nevezzük.

Határozzuk meg a (13), (14) explicit Euler-módszer lokális approximációs hibáját és konzisztenciarendjét!

$$\Psi_n^{(1)} = -\frac{u(t_{n+1}) - u(t_n)}{\tau} + f(t_n, u(t_n)) \tag{17}$$

Meg kell mutatnunk, hogy (17) nullához fog tartani, valamint azt is, hogy milyen rendben. Tegyük fel, hogy $u(t)$ kétszer folytonosan differenciálható! Taylor-sorba fejtjük $u(t_{n+1})$-et a t_n pont körül:

$$u(t_{n+1}) = u(t_n + \tau) = u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2!}u''(t_n) + \mathcal{O}(\tau^3). \tag{18}$$

Másrészt

$$f(t_n, u(t_n)) = u'(t_n). \tag{19}$$

A (18), (19) egyenleteket behelyettesítjük a (15) képletbe:

$$\Psi_n^{(1)} = -\frac{(u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2!}u''(t_n) + \mathcal{O}(\tau^3)) - u(t_n)}{\tau} + u'(t_n) =$$
$$= -\frac{\tau}{2!}u''(t_n) + \mathcal{O}(\tau)^2 \equiv \mathcal{O}(\tau).$$

Tehát az explicit Euler-módszer elsőrendben konzisztens.

3.3. Az explicit Euler-módszer konvergenciája

Még mindig kérdés, hogy az $y_n \approx u(t_n)$ közelítés teljesül-e. Ahhoz, hogy ez teljesüljön, a korábban definiált konvergencia szükséges. Hogyan tudjuk garantálni ezt a konvergenciát? Vizsgáljuk meg, hogy a (7) tulajdonság teljesül-e a (13), (14) módon előállított y_n rácsfüggvényekre! Tehát a kérdés az, hogy a globális hiba nullához fog-e tartani. Ha a válasz igen, akkor konvergensnek mondhatjuk a módszert, vagyis teljesül rá az alábbi feltétel:

$$\lim_{\tau \to 0} |e_n| = 0 \Longleftrightarrow \lim_{\tau \to 0} e_n = 0.$$

Az előző fejezetben definiáltak szerint

$$\frac{e_{n+1} - e_n}{\tau} = \Psi_n^{(1)} + \Psi_n^{(2)}, \tag{20}$$

ahol $\Psi_n^{(1)}$ és $\Psi_n^{(2)}$ (15), (16) szerint adhatók meg. Adjunk becslést a $|\Psi_n^{(1)}|$ kifejezésre! A (18) képlet alapján látható, hogy amennyiben u kétszer folytonosan differenciálható

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

ahol $\xi_n \in [t_n, t_{n+1}]$ egy adott pont. Mivel $f(t_n, u(t_n)) = u'(t_n)$, valamint $\Psi_n^{(1)}$ (15) szerinti, ezért $|\Psi_n^{(1)}| \leq \frac{M_2}{2}\tau$, ahol

$$M_2 := \max_{[0,t^*]} |u''(t)|. \tag{21}$$

Az f függvény Lipschitz-tulajdonsága miatt

$$|\Psi_n^{(2)}| \leq L|(u_n + e_n) - (u_n)| = L \cdot |e_n|. \tag{22}$$

Ekkor a (20), (22) egyenletekből az alábbi kifejezésre jutunk:

$$e_{n+1} = e_n + \tau(\Psi_n^{(1)} + \Psi_n^{(2)}).$$

Tehát:

$$|e_{n+1}| \leq |e_n| + \tau(|\Psi_n^{(1)}| + |\Psi_n^{(2)}|) \leq |e_n| + \tau|\Psi_n^{(1)}| + \tau L \cdot |e_n| = (1+\tau L)|e_n| + \tau|\Psi_n^{(1)}|.$$

$$\begin{aligned}
|e_n| &\leq (1+\tau L)|e_{n-1}| + \tau|\Psi_{n-1}^{(1)}| \leq \\
&\leq (1+\tau L)\Big((1+\tau L)|e_{n-2}| + \tau|\Psi_{n-2}^{(1)}|\Big) + \tau|\Psi_{n-1}^{(1)}| = \\
&= (1+\tau L)^2|e_{n-2}| + \tau\Big(|\Psi_{n-2}^{(1)}|(1+\tau L) + |\Psi_{n-1}^{(1)}|\Big) \leq \dots \\
&\dots \leq (1+\tau L)^n|e_0| + \tau\sum_{j=0}^{n-1}(1+\tau L)^j|\Psi_{n-1-j}^{(1)}| \leq \\
&\leq (1+\tau L)^n\Big(|e_0| + \tau\sum_{j=0}^{n-1}|\Psi_{n-1-j}^{(1)}|\Big)
\end{aligned} \tag{23}$$

Az utolsó lépésben az $(1+\tau L)^j < (1+\tau L)^n, j = 0, 1, \dots, n-1$ egyenlőtlenséget alkalmaztuk. Mivel $1 + x \leq e^x$ $\forall x \geq 0$ esetén, valamint $n\tau = t^*$, ezért

$$(1+\tau L)^n \leq (e^{\tau L})^n = e^{n\tau L} = e^{t^* L}. \tag{24}$$

Így (23)-ba behelyettesítve (24)-et és (21)-et a következőt kapjuk:

$$\begin{aligned}
|e_n| &\leq e^{t^* L}\Big(|e_0| + \tau\frac{M_2}{2}\tau n\Big) \\
|e_n| &\leq e^{t^* L}\Big(|e_0| + \frac{M_2 t^*}{2}\tau\Big), \ \forall n = 1, 2, \dots, N.
\end{aligned} \tag{25}$$

Mivel $e_0 = y_0 - u(0) = u_0 - u_0 = 0$, így (25)-ből

$$\begin{aligned}
|e_n| &\leq K \cdot \tau \\
K &= e^{t^* L} \cdot \frac{M_2}{2} t^*
\end{aligned} \tag{26}$$

következik. A (26) becslés azt mutatja meg, hogy megfelelően finomodó ω_τ rácshálókon $\tau \to 0$ esetén $e_n \to 0$, és emellett $e_n = \mathcal{O}(\tau)$. Tehát az explicit Euler-módszer elsőrendben konvergens.

3.4. Az implicit Euler-módszer

Az explicit Euler-módszernél tárgyalt rácsfüggvényt máshogy is meghatározhatjuk. Abban az esetben a deriváltat a jobb oldali véges differenciahányadossal közelítettük, és a (13) képletet kaptuk, de vegyük észre, hogy a (13) bal oldalán szereplő hányados egyben a megoldásfüggvény t_{n+1}-beli deriváltjának bal oldali véges differenciahányadossal való közelítése is. Ezért a jobb oldalt kicserélhetjük az $f(t_{n+1}, y_{n+1})$ értékre.

$$f(t_{n+1}, u(t_{n+1})) = u'(t_{n+1}) \approx y'(t_{n+1}) \approx \frac{y(t_{n+1}) - y(t_n)}{\tau} = f(t_{n+1}, y(t_{n+1}))$$

Az

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1}), \ n = 0, 1, \ldots, N-1 \tag{27}$$

$$y_0 = u_0 \tag{28}$$

módon definiált módszert implicit Euler-módszernek nevezzük.

3.2. Megjegyzés. A (27), (28) implicit Euler-módszert azért nevezzük implicitnek, mert y_n ismeretében az y_{n+1} értékét minden egyes lépésben egy (általában nemlineáris) egyenlet megoldásával tudjuk csak meghatározni.

3.5. Az implicit Euler-módszer vizsgálata

Határozzuk meg az implicit Euler-módszer lokális approximációs hibáját és konzisztenciarendjét! Az $y_n = e_n + u_n$ képletet behelyettesítjük a (27) rekurzióba:

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

Ezt átrendezve, majd hozzáadva és kivonva belőle az $f(t_{n+1}, u_{n+1})$ értéket az alábbi kifejezést kapjuk:

$$\frac{e_{n+1} - e_n}{\tau} = \left(-\frac{u_{n+1} - u_n}{\tau} + f(t_{n+1}, u_{n+1}) \right) \left(-f(t_{n+1}, u_{n+1}) + f(t_{n+1}, e_{n+1} + u_{n+1}) \right).$$

A korábbi $\Psi_n^{(1)}$ jelöléssel az implicit Euler-módszer lokális approximációs hibája:

$$\Psi_n^{(1)} = -\frac{u_{n+1} - u_n}{\tau} + f(t_{n+1}, u_{n+1}) = -\frac{u(t_{n+1}) - u(t_n)}{\tau} + f(t_{n+1}, u(t_{n+1})). \tag{29}$$

Meg kell mutatnunk, hogy (29) nullához fog tartani, valamint azt is, hogy milyen rendben. Tegyük fel, hogy $u(t)$ kétszer folytonosan differenciálható! Fejtsük Taylor-sorba $u(t_n)$-t a t_{n+1} pont körül:

$$u(t_n) = u(t_{n+1} - \tau) = u(t_{n+1}) - \tau u'(t_{n+1}) + \frac{\tau^2}{2!} u''(t_{n+1}) + \mathcal{O}(\tau^3). \tag{30}$$

Másrészt

$$f(t_{n+1}, u(t_{n+1})) = u'(t_{n+1}). \tag{31}$$

A (30), (31) egyenleteket behelyettesítjük a (29) képletbe:

$$\Psi_n^{(1)} = -\frac{u(t_{n+1}) - \big(u(t_{n+1}) - \tau u'(t_{n+1}) + \frac{\tau^2}{2!} u''(t_{n+1}) + \mathcal{O}(\tau^3)\big)}{\tau} + u'(t_{n+1}) =$$
$$= \frac{\tau}{2!} u''(t_{n+1}) + \mathcal{O}(\tau)^2 \equiv \mathcal{O}(\tau).$$

Tehát az implicit Euler-módszer elsőrendben konzisztens.

3.6. A trapézmódszer vizsgálata

Az egylépéses módszerek közül fontos még megemlítenünk a trapézmódszert, amely a következő alakban írható fel:

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{2}(f(t_n, y_n) + f(t_{n+1}, y_{n+1})), \ n = 0, 1, \ldots, N-1 \tag{32}$$
$$y_0 = u_0.$$

A trapézmódszer lokális approximációs hibája a következőképpen írható fel:

$$\Psi_n^{(1)} = -\frac{u_{n+1} - u_n}{\tau} + \frac{1}{2}\big(f(t_n, u_n) + f(t_{n+1}, u_{n+1})\big) =$$
$$= -\frac{u(t_{n+1}) - u(t_n)}{\tau} + \frac{1}{2}\big(f(t_n, u(t_n)) + f(t_{n+1}, u(t_{n+1}))\big).$$

A (9) képletet felhasználva:

$$\Psi_n^{(1)} = -\frac{u(t_{n+1}) - u(t_n)}{\tau} + \frac{1}{2}\big(u'(t_n) + u'(t_{n+1})\big). \tag{33}$$

A t_n pont körül fejtsük Taylor-sorba $u(t_{n+1})$-et és $u'(t_{n+1})$-et:

$$u(t_{n+1}) = u(t_n + \tau) = u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2!} u''(t_n) + \mathcal{O}(\tau^3), \tag{34}$$

$$u'(t_{n+1}) = u'(t_n + \tau) = u'(t_n) + \tau u''(t_n) + \mathcal{O}(\tau^2). \tag{35}$$

A (34), (35)-öt behelyettesítve a (33) képletbe a:

$$\Psi_n^{(1)} = -\frac{(u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2!}u''(t_n) + \mathcal{O}(\tau^3)) - u(t_n)}{\tau} + \frac{1}{2}\big(u'(t_n) + u'(t_n) + \tau u''(t_n) + \mathcal{O}(\tau^2)\big)$$

összefüggést kapjuk. Ezt egyszerűsítve: $\Psi_n^{(1)} = \mathcal{O}(\tau^2)$, tehát a trapézmódszer másodrendben konzisztens.

3.7. A θ-módszer

Az eddig ismertetett módszerek a következő, úgynevezett θ-módszer speciális esetei:

$$\begin{aligned} \frac{y_{n+1} - y_n}{\tau} &= (1-\theta)f(t_n, y_n) + \theta f(t_{n+1}, y_{n+1}), \ n = 0, 1, \ldots, N-1 \\ y_0 &= u_0. \end{aligned} \tag{36}$$

A $\theta \in [0, 1]$ rögzített paraméter különböző értékeire különböző módszereket kapunk:

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

3.8. Példa

Tekintsük a következő kezdetiérték-feladatot:

$$u'(t) = -u(t) + t + 1, \tag{37}$$

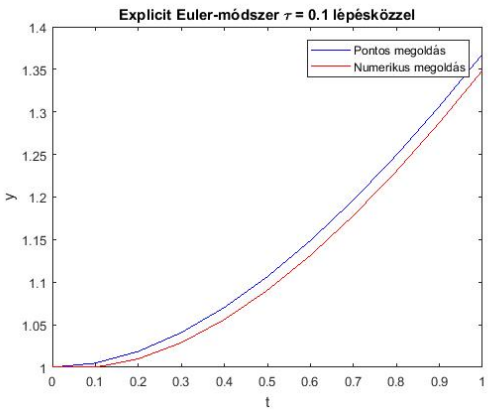
ahol $t \in [0, 1]$ és $u(0) = 1$ a kezdeti feltétel.

Ennek a feladatnak a pontos megoldása a következő:

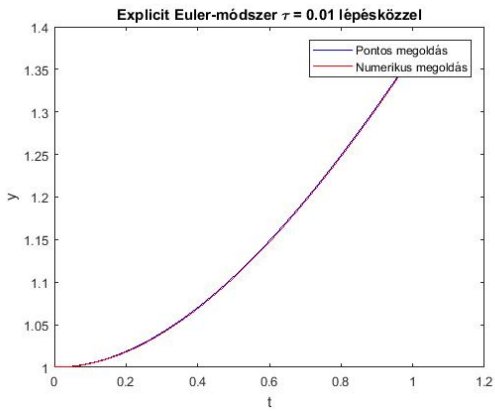
$$u(t) = e^{-t} + t.$$

Ezt a megoldást szeretnénk összehasonlítani az explicit, valamint az implicit Euler-módszerrel kapott megoldásokkal különböző lépésközöket alkalmazva.

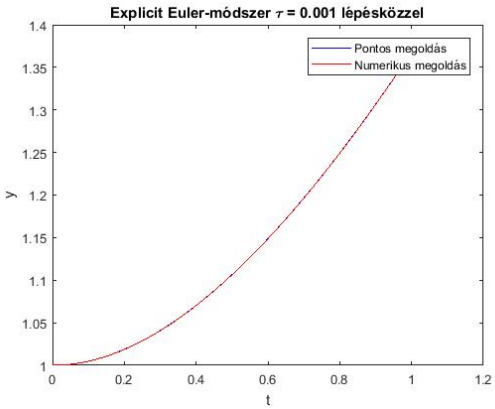
Az 1. és 2. ábrákon látszik, hogy a τ lépésköz csökkentésével, azaz egyre finomodó rácsháló választása esetén a numerikus megoldás grafikonja közeledik a pontos megoldás grafikonjához.



(a) A $\tau = 0.1$ lépésközű EEM.

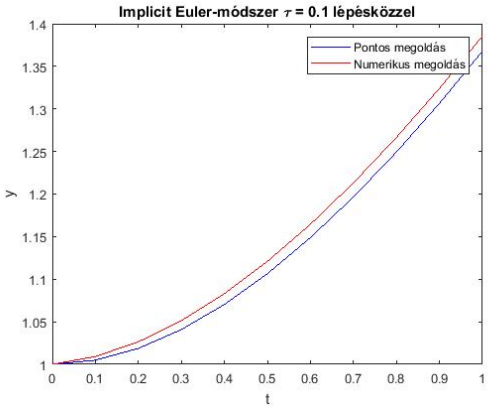


(b) A $\tau = 0.01$ lépésközű EEM.

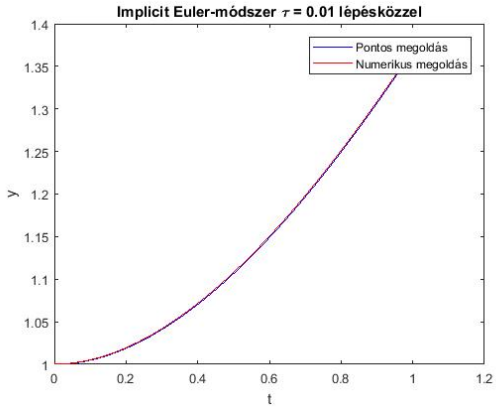


(c) A $\tau = 0.001$ lépésközű EEM.

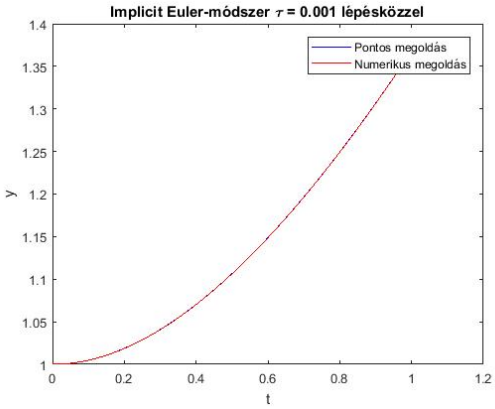
1. ábra. Az explicit Euler-módszerrel kapott numerikus megoldás különböző lépésközök esetén.



(a) A $\tau = 0.1$ lépésközű IEM.



(b) A $\tau = 0.01$ lépésközű IEM.



(c) A $\tau = 0.001$ lépésközű IEM.

2. ábra. Az implicit Euler-módszerrel kapott numerikus megoldás különböző lépésközök esetén.

Ezt a fejezetet az [1], [4], [5] jegyzetek segítségével dolgoztuk fel.

4. A numerikus módszerek stabilitása

Ebben a fejezetben a numerikus módszerek stabilitásáról lesz szó. A numerikus módszer konvergenciájához nem elég a konzisztenciarendjét ismernünk, szükségünk van még arra is, hogy a módszer 0-stabil legyen.

4.1. Tétel. *Ha egy numerikus módszer p-ed rendben konzisztens és 0-stabil, akkor p-ed rendben konvergens.*

A 0-stabilitás lényegében azt jelenti, hogy ha egy kicsit változtatunk a kezdeti feltételen, akkor a numerikus megoldások is csak kicsit térnek el egymástól az egész időintervallumon.

4.1. Definíció. Egy numerikus módszer 0-stabil, ha $\exists K > 0$, amelyre teljesül az

$$|e_i| \leq K\left(|e_0| + \sum_{j=0}^{i-1} |d_j|\tau\right)$$

reláció minden $i = 1, \ldots, N$ esetén.

A definícióban szereplő e_i jelöli a globális hibát, amely az összes lépés alatt keletkezik, d_i pedig a lokális hibát, vagyis azt a hibát, amely egy lépés alatt keletkezik.

Az alkalmazások során fontos a rögzített rácson való stabilitás vizsgálata is. Ezt a következő, úgynevezett *Dahlquist-féle tesztfeladat*on szokás elemezni:

$$y'(t) = \lambda y(t) \tag{38}$$

$$y(0) = y_0, \tag{39}$$

ahol $\lambda \in \mathbb{C}$, $\mathrm{Re}(\lambda) \leq 0$, $y_0 \in \mathbb{R}$, $t \geq 0$.

Ennek a feladatnak a pontos megoldása: $y(t) = y_0 e^{\lambda t}$. Látszik, hogy ez a függvény $\mathrm{Re}(\lambda) \leq 0$ esetén a $[0, \infty)$ intervallumon korlátos, továbbá $\mathrm{Re}(\lambda) < 0$ esetén nullához tart, amidőn $t \to \infty$. A numerikus megoldástól elvárjuk, hogy szintén teljesítse ezeket a tulajdonságokat. Vagyis azt mondhatjuk, hogy egy rögzített lépésközű rácshálón vett numerikus megoldás stabil, ha a (38), (39) tesztfeladatra alkalmazva a numerikus megoldás korlátos marad, miközben $t \to \infty$.

Legyen $z := \lambda\tau$. Ha a (38), (39) tesztfeladatra alkalmazzuk a numerikus módszert, akkor az n-edik lépés a következőképpen írható fel:

$$y_{n+1} = R(z)y_n, \tag{40}$$

ahol az $R(z)$ függvény a módszer stabilitási függvénye.

4.2. Definíció. Egy numerikus módszer stabilitási tartománya a $\{z \in \mathbb{C} : |R(z)| \leq 1\}$ számhalmaz.

4.3. Definíció. Egy módszert A-stabilnak nevezünk, ha a bal oldali komplex félsík része a stabilitási tartományának.

4.4. Definíció. Egy numerikus módszert erősen A-stabilnak nevezünk, ha A-stabil, és

$$\lim_{|z|\to\infty} |R(z)| < 1.$$

4.5. Definíció. Egy numerikus módszert L-stabilnak nevezünk, ha A-stabil, és

$$\lim_{|z|\to\infty} R(z) = 0.$$

4.1. Megjegyzés. Minden L-stabil módszer erősen A-stabil, és minden erősen A-stabil módszer A-stabil.

A következő alfejezetekben az explicit és az implicit Euler-módszer stabilitási tartományával foglalkozunk.

4.1. Az explicit Euler-módszer stabilitási tartománya

Ha az explicit Euler-módszert a (38), (39) Dahlquist-féle tesztfeladatra alkalmazzuk, akkor a következőt kapjuk:

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

(40) alapján látszik, hogy az explicit Euler-módszer stabilitási függvénye: $R(z) = 1 + z$. Tehát a stabilitási tartománya azon z komplex számok halmaza, melyekre $|R(z)| \leq 1$, azaz $|1 + z| \leq 1$. Vizsgáljuk meg, mit jelent ez egy valós $\lambda \leq 0$ szám esetén a választható lépésköz hosszára nézve!

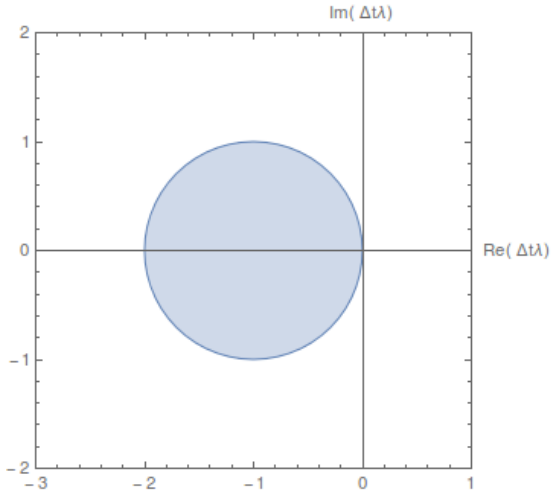
$$|1 + z| \leq 1 \;\Leftrightarrow\; |1 + \lambda\tau| \leq 1 \;\Leftrightarrow\; -1 \leq \lambda\tau \leq 1 \;\Leftrightarrow\; -2 \leq \lambda\tau \leq 0.$$

Mivel tudjuk, hogy $\lambda \leq 0$, valamint a τ lépésköz mindig pozitív, így a $\lambda\tau \leq 0$ lesz, tehát a stabilitási feltétel: $-2 \leq \lambda\tau$, azaz

$$\tau \leq \frac{-2}{\lambda} = \frac{2}{|\lambda|}.$$

Ez azt jelenti, hogy ha ennél hosszabb τ lépésközt választunk, akkor a (38), (39) feladatra alkalmazott explicit Euler-módszer megoldása instabillá válik: a numerikus megoldás elszáll a végtelenbe, ahogy időben előrefelé lépegetünk.

Az 3. ábrán az explicit Euler-módszer teljes stabilitási tartománya látható. A stabilitási tartomány nem foglalja magába a bal oldali félsíkot, tehát az explicit Euler-módszer nem A-stabil, ebből következően nem erősen A-stabil, és nem L-stabil.



3. ábra. Az explicit Euler-módszer stabilitási tartománya.

4.2. Az implicit Euler-módszer stabilitási tartománya

Ha az implicit Euler-módszert a (38), (39) Dahlquist-féle tesztfeladatra alkalmazzuk, akkor a következőt kapjuk:

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

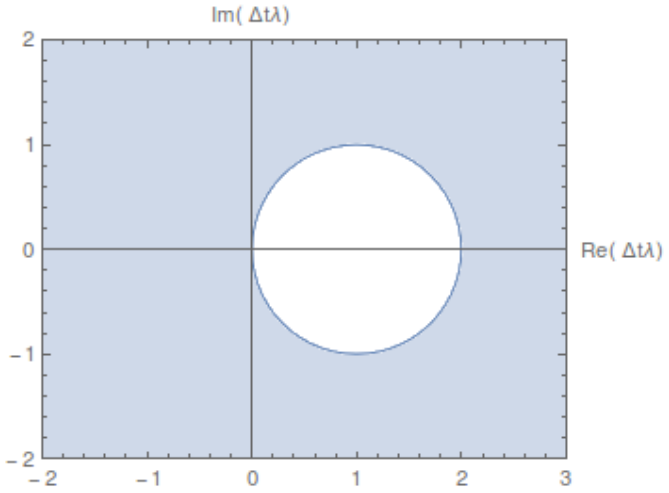
$$y_{n+1}(1 - \tau\lambda) = y_n$$

$$y_{n+1} = \frac{1}{1-z} y_n.$$

(40) alapján látszik, hogy az implicit Euler-módszer stabilitási függvénye: $R(z) = \frac{1}{1-z}$. Tehát a stabilitási tartománya azon z komplex számok halmaza, melyekre $|R(z)| \leq 1$, azaz $\left|\frac{1}{1-z}\right| \leq 1$. Vizsgáljuk meg, mit jelent ez egy valós $\lambda \leq 0$ szám esetén a választható lépésköz hosszára nézve!

$$\left|\frac{1}{1-z}\right| \leq 1 \;\Leftrightarrow\; \left|\frac{1}{1-\lambda\tau}\right| \leq 1 \;\Leftrightarrow\; -1 \leq \frac{1}{1-\lambda\tau} \leq 1.$$

Mivel tudjuk, hogy $\lambda \leq 0$, valamint a τ lépésköz mindig pozitív, így a $\lambda\tau \leq 0$ lesz. Ezt a szorzatot a tört nevezőjében 1-ből vonjuk ki, tehát így egy 1-nél nagyobb vagy egyenlő számot kapunk, és az 1-et leosztva ezzel, a tört értékére mindig igaz a fenti reláció. Tehát a módszer bármilyen τ lépésköz esetén stabil megoldást ad.



4. ábra. Az implicit Euler-módszer stabilitási tartománya.

Az 4. ábrán az implicit Euler-módszer teljes stabilitási tartománya látható. A stabilitási tartomány magába foglalja a bal oldali félsíkot, tehát az implicit Euler-módszer A-stabil. A stabilitási függvény alakjából az is látható, hogy ha $t \to \infty$, akkor a stabilitási függvény értéke nullához konvergál. Tehát az implicit Euler-módszer erősen A-stabil, sőt L-stabil.

Ezt a fejezetet a [3], [6], [8], [9] jegyzetek segítségével dolgoztuk fel.

5. A Richardson-extrapoláció

5.1. Bevezetés

A korábbi fejezetekben szó volt a közönséges differenciálegyenletekre vonatkozó klasszikus kezdetiérték-feladatról, amit a következőképpen definiáltunk:

$$u'(t) = f(t, u(t)) \tag{41}$$

$$u(t_0) = u_0, \tag{42}$$

ahol

- $t \in [t_0, T],\ t_0 < T,\ u \in \mathbb{R}^d,\ d \geq 1,\ f : G \to \mathbb{R}^d$, ahol $G \subset \mathbb{R}^{d+1}$,
- t az ismeretlen változó,
- d az egyenletek számát adja meg,
- f pedig egy adott folytonos leképezés,
- $u(t)$ egy d-dimenziós vektor, amely a t változótól függ és az ismeretlen függvényt reprezentálja. Az egyszerűség kedvéért feltesszük, hogy $d = 1$. Ekkor tehát a keresett függvény $u : [t_0, T] \to \mathbb{R}$ típusú.

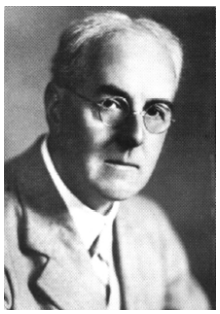
Továbbá azt is feltesszük, hogy létezik megoldása a feladatnak, és az egyértelmű. Ennek feltételeit korábban már tárgyaltuk (Lipschitz-tulajdonság). Azt is láttuk, hogy általában a (41), (42) módon definiált feladatnak nem tudjuk kiszámítani a pontos megoldását, ezért közelítő eljárásokat alkalmazunk. Bevezettük az $\omega_\tau := \{t_n = n \cdot \tau, n = 0, 1, \ldots\}$ ekvidisztáns felosztású rácshálót, ahol $\tau > 0$ az időbeli lépésköz volt.

A numerikus modellezés során gyakran felmerülő probléma, hogy adott lépésköz esetén a numerikus megoldás nem kellően pontos. Ilyen helyzetekben az a megszokott eljárás, hogy az így kapott eredményt figyelmen kívül hagyjuk, majd újra kezdjük a számolást, kisebb lépésközt alkalmazva. Ez azonban a sok felesleges számolás miatt igen költséges.

Ebben a fejezetben a Richardson-extrapolációval foglalkozunk, amely nagyon népszerű numerikus eljárás. Segítségével hatékonyan növelhető egy időintegrálási módszer pontossága. Ezt a számítási eszközt különféle numerikus módszerekkel kombinálva használhatjuk.

Lényege, hogy ugyanazon közelítő módszert két különböző lépésközzel alkalmazzuk, és ezeket megfelelő módon kombinálva a p-ed rendű módszerből $(p+1)$-ed rendben pontos eredményt nyerhetünk. Két változata ismert: az aktív és a passzív Richardson-extrapoláció.

Az eredeti ötlet Lewis Fry Richardson angol matematikustól származik [10, 11], aki többek között az időjárás előrejelzésének problémájával is foglalkozott. Matematikailag modellezte és kidolgozta a légköri áramlatok folyamatait. Bár első próbálkozása nem volt sikeres, nem kapott pontos eredményt, munkássága rendkívül fontos a mai időjárás-előrejelzés szempontjából.



5. ábra. Lewis Fry Richardson.

Tekintsük a (41), (42) klasszikus kezdetiérték-feladatot. Tegyük fel, hogy a közönséges differenciálegyenlet közelítésére valamely korábban tárgyalt numerikus módszert alkalmazzuk lépésről lépésre. Jelölje $u(t_n) = u_n$ a feladat pontos megoldását az ω_τ rácsháló valamely t_n pontjában. Ezen feltevések mellett a Richardson-extrapoláció az alábbiak szerint írható le.

Ha a számításokat már minden t_i $(i = 0, 1, \ldots, n-1)$ rácspontra végrehajtottuk egy p-ed rendben konvergens numerikus módszerrel, akkor a következő három lépéssel adható meg a következő, t_n rácspontbeli közelítő érték $(y_\tau(t_n) = y(t_n) = y_n)$:

1. τ lépésközzel számoljuk ki $u(t_n)$ approximációját, jelölje ezt a közelítést z_n,

2. 0.5τ lépésközzel (azaz két lépés megtételével) számoljuk ki $u(t_n)$ approximációját, jelölje ezt a közelítést w_n,

3. az alábbi képletet alkalmazva kiszámítjuk az y_n közelítő megoldást:

$$y_n = \frac{2^p w_n - z_n}{2^p - 1}. \tag{43}$$

A fenti három lépéssel definiált algoritmust nevezzük Richardson-extrapolációnak. Ezt a módszert L. F. Richardson mutatta be és dolgozta ki 1911-ben és 1927-ben [10, 11]. Ne feledjük, hogy az ötlet valójában nagyon általános. A fenti algoritmus a közönséges differenciálegyenletek bármely egylépéses numerikus módszerrel történő megoldására alkalmazható. A használatához csak az alábbi két követelményt kell kielégíteni:

- z_n és w_n kiszámításához ugyanazt az egylépéses módszert kell alkalmazni,
- ha a kiválasztott módszer rendje p, akkor ezen p értéket kell behelyettesíteni a (43) képletbe, tehát ez határozza meg a lineáris kombinációban szereplő súlyokat.

5.2. A Richardson-extrapoláció pontossági rendje

Tegyük fel, hogy z_n és w_n a korábbiak szerint lett kiszámolva egy p-ed rendben konvergens numerikus módszerrel. Továbbá ha feltesszük, hogy a közönséges differenciálegyenlet pontos $u(t)$ megoldása megfelelően sima (azaz $u(t)$ $(p+1)$-szer folytonosan differenciálható függvény, ami erősebb megkötés a korábban említett Lipschitz-tulajdonságnál), akkor a következő két egyenlet írható fel ezen értékekre:

$$u(t_n) - z_n = \tau^p K + \mathcal{O}(\tau^{p+1}), \tag{44}$$

$$u(t_n) - w_n = (0.5\tau)^p K + \mathcal{O}(\tau^{p+1}). \tag{45}$$

A (44), (45) képletben szereplő K mennyiség függ a feladattól és a z_n és w_n számítása során alkalmazott numerikus módszertől, de nem függ az alkalmazott lépésköztől. A K tag kiejtésével a fenti két egyenletből az

$$u(t_n) - \frac{2^p w_n - z_n}{2^p - 1} = \mathcal{O}(\tau^{p+1}) \tag{46}$$

egyenlőség adódik.

Vegyük észre, hogy a (46) képlet bal oldalán szereplő második kifejezés pontosan a Richardson-extrapoláció alkalmazásával kapott y_n közelítés ((43) képlet). A fentiek alapján azonnal megkaphatjuk a következő összefüggést:

$$u(t_n) - y_n = \mathcal{O}(\tau^{p+1}). \tag{47}$$

Összehasonlítva az így kapott (47) képletet a (44) és (45) képletekkel, egyértelműen látszik, hogy az y_n $(p+1)$-ed rendben közelíti a megoldást, így megfelelően kicsi lépésközökre mind a z_n, mind a w_n közelítésnél pontosabb. Ez azt jelenti, hogy a Richardson-extrapolációt felhasználhatjuk a numerikus megoldás pontosságának növelésére.

5.3. A pontosság további növelése

Az előzőekben láttuk, hogy a Richardson-extrapolációval hatékonyan tudjuk növelni a numerikus módszerek pontosságát, de vajon tudjuk-e tovább növelni ezt a pontosságot? Azaz megadható-e olyan általános számítási eljárás, amelynek segítségével egy p-ed rendben pontos módszerből $(p+2)$-ed rendben pontos módszert nyerünk? Vegyük alapul az előző alfejezetben használt (44) és (45) képleteket. Módosítsuk a kiszámolt közelítésekre használt jelöléseket, és a jobb oldalon vegyünk figyelembe egy további tagot a sorfejtésből:

$$u(t_n) - z_n^{[1]} = \tau^p K + \tau^{p+1} L + \mathcal{O}(\tau^{p+2}), \tag{48}$$

$$u(t_n) - z_n^{[2]} = (0.5\tau)^p K + (0.5\tau)^{p+1} L + \mathcal{O}(\tau^{p+2}). \tag{49}$$

Tegyük fel, hogy egy harmadik $z_n^{[3]}$ közelítést is meghatároztunk 0.25τ lépésközzel:

$$u(t_n) - z_n^{[3]} = (0.25\tau)^p K + (0.25\tau)^{p+1} L + \mathcal{O}(\tau^{p+2}). \tag{50}$$

A K és L együtthatókat szeretnénk kiküszöbölni a fenti egyenlőségekből. Először a K paramétert tudjuk eliminálni: (49)-et megszorozzuk 2^p-nel, majd az így kapott képletből kivonjuk (48)-at. Hasonlóképpen szorozzuk meg (50)-et $4^p = 2^{2p}$-nel, majd ebből vonjuk ki (48)-at. Ennek eredménye:

$$(2^p - 1)u(t_n) - 2^p z_n^{[2]} + z_n^{[1]} = -0.5\tau^{p+1} L + \mathcal{O}(\tau^{p+2}) \tag{51}$$

$$(2^{2p} - 1)u(t_n) - 2^{2p} z_n^{[3]} + z_n^{[1]} = -0.75(\tau)^{p+1} L + \mathcal{O}(\tau^{p+2}). \tag{52}$$

Az L konstanst a következőképpen tudjuk eliminálni: (51)-et megszorozzuk -0.75-dal és (52)-t megszorozzuk 0.5-del:

$$-0.75(2^p - 1)u(t_n) + (0.75)2^p z_n^{[2]} - 0.75 z_n^{[1]} = (0.75)0.5\tau^{p+1} L + \mathcal{O}(\tau^{p+2}) \tag{53}$$

$$0.5(2^{2p} - 1)u(t_n) - (0.5)2^{2p} z_n^{[3]} + 0.5 z_n^{[1]} = 0.5(-0.75)(\tau)^{p+1} L + \mathcal{O}(\tau^{p+2}). \tag{54}$$

Az (53) és (54) egyenlőségek összeadásával:

$$\left(2^{2p-1} - 3(2^{p-2}) + 0.25\right)u(t_n) - 2^{2p-1}z_n^{[3]} + 3(2^{p-2})z_n^{[2]} - 0.25z_n^{[1]} = \mathcal{O}(\tau^{p+2}). \tag{55}$$

Az (55) formula a következő alakra hozható:

$$u(t_n) = \frac{2^{2p-1}z_n^{[3]} - 3(2^{p-2})z_n^{[2]} + 0.25z_n^{[1]}}{2^{2p-1} - 3(2^{p-2}) + 0.25} + \mathcal{O}(\tau^{p+2}). \tag{56}$$

Legyen

$$u_n = \frac{2^{2p-1}z_n^{[3]} - 3(2^{p-2})z_n^{[2]} + 0.25z_n^{[1]}}{2^{2p-1} - 3(2^{p-2}) + 0.25}. \tag{57}$$

Ekkor (56) átírható a következő alakra:

$$u(t_n) = u_n + \mathcal{O}(\tau^{p+2}). \tag{58}$$

Ebből világosan látszik, hogy u_n a maga $(p+2)$-ed rendű pontosságával két renddel pontosabb $z_n^{[1]}, z_n^{[2]}, z_n^{[3]}$ közelítések mindegyikénél, amelyek mind p-ed rendben pontosak.

5.1. Megjegyzés. Fontos megemlíteni, hogy ezen algoritmus számítási költsége elég magas, mivel szükséges hozzá elvégezni:

1. egy τ lépésközzel számított approximációt: $z_n^{[1]}$,
2. egy 0.5τ lépésközzel számított approximációt: $z_n^{[2]}$, és
3. egy 0.25τ lépésközzel számított approximációt: $z_n^{[3]}$.

Cserébe azonban hosszabb τ lépésközt választhatunk ugyanazon pontosság eléréséhez. Ezt a módszert ismételt Richardson-extrapolációnak nevezzük [3].

5.4. Hibabecslés és a lépésköz automatikus szabályozása

Vannak olyan esetek, amikor nem ekvidisztáns felosztású rácshálón keressük a közelítő megoldást, hanem egy változó $\tau_n > 0$ lépésközű $\omega_{\tau'} := \{t_n = t_{n+1} + \tau_n,\ n = 1, 2, \ldots\}$ rácshálón. Ez a τ_n lépésköz különbözhet a τ_{n-1} lépésköztől, amit az $(n-1)$-edik lépésben használtunk, valamint τ_{n+1}-től is, amit az $(n+1)$-edik lépésben használunk. Azonban nem lehet tetszés szerint váltogatni ezeket a lépésközöket, bizonyos szabályoknak teljesülniük kell:

- az így számolt közelítéseknek pontosnak kell maradniuk,
- biztosítani kell a 0-stabilitást, valamint
- a műveletigényt hatékonyan kell csökkenteni.

A változó lépésköz használata gyakran növeli a számítási folyamat hatékonyságát, így akár egy optimális lépésközhosszúságot is be lehet állítani.

A Richardson-extrapolációt használhatjuk, és valójában nagyon gyakran használják oly módon, hogy megbecsülik a közelítő megoldás számításakor fellépő hiba vezető tagját, majd ezután meghatározzák az optimális lépésközt, amely remélhetőleg hatékonyan alkalmazható a következő lépésben. Vegyük figyelembe, hogy a (44) és (45) kifejezések nem használhatók közvetlenül a hibabecslésnél, mert K értéke általában nem ismert. Ez azt jelenti, hogy ezt a paramétert eliminálni kell annak érdekében, hogy meg lehessen becsülni a hibát, majd azt követően egy optimális lépésközt lehessen meghatározni.

Ha a (44) egyenletből a (45) egyenletet kivonjuk, az alábbi kifejezést kapjuk K-ra:

$$K = \frac{2^p(w_n - z_n)}{\tau^p(2^p - 1)} + \mathcal{O}(\tau^{p+1}). \tag{59}$$

Az így kapott (59) képletet behelyettesítve a (45) képletbe az

$$u(t_n) - w_n = \frac{w_n - z_n}{2^p - 1} + \mathcal{O}(\tau^{p+1}) \tag{60}$$

egyenlőség adódik, ami azt mutatja, hogy az n-edik időpillanatban a

$$HIBA_n = \left| \frac{w_n - z_n}{2^p - 1} \right| \tag{61}$$

mennyiség alkalmas a w_n közelítő megoldás számítása során fellépő globális hiba vezető tagjának becslésére, ha a τ lépésköz megfelelően kicsi. A

$$\tau_{\text{új}} = \omega \sqrt[p]{\frac{TOL}{HIBA_n}} \tau \tag{62}$$

formula segítségével automatikusan kontrollálhatjuk a lépésközt. Az (62)-beli TOL paraméter értékét mi határozzuk meg, a (41), (42) feladat megoldásának a kívánt pontosságát adjuk meg vele. Ha a (61) alapján számolt globális hiba jelentősen eltér TOL értékétől, akkor $HIBA_n$ felhasználható egy új ($\tau_{\text{új}}$) lépésköz meghatározására. Az (62) képletben szereplő ω egy biztonsági paraméter, értéke 0 és 1 között változhat, általában $\omega = 0.9$. Tehát a Richardson-extrapoláció alkalmazható közönséges differenciálegyenlet-rendszereket megoldó programok lépésközének automatikus szabályozására is.

5.5. Az aktív és a passzív Richardson-extrapoláció

A (43) formula valójában csak azt határozza meg, hogyan számolhatjuk ki az y_n extrapolált közelítést minden egyes lépésben, ahol $n = 0, 1, 2, \ldots, N$, azt feltételezve, hogy a két közelítés: z_n és w_n rendelkezésünkre áll. Ez a képlet önmagában még nem teljesen írja le, hogyan működik a Richardson-extrapoláció algoritmusa. Az algoritmus akkor lesz jól definiálva, ha pontosítjuk, hogy mi történik az y_n ($n = 0, 1, 2, \ldots, N$) közelítésekkel. Két lehetéges választás van:

1. egy adott n esetén a kiszámolt y_n közelítés nem szerepel a további számításokban (el lehet tárolni a későbbi fehasználás céljából) (passzív Richardson-extrapoláció),

2. egy adott n esetén a kiszámolt y_n közelítést közvetlenül felhasználjuk y_{n+1} kiszámításához (aktív Richardson-extrapoláció).

Tehát a Richardson-extrapolációnak két különböző változatát kapjuk: az aktív és a passzív Richardson-extrapolációt. Világos, hogy miért megfelelő jelző a passzív szó az első változatra: a numerikus megoldást minden időpillanatban a két különböző lépésközű megoldás kombinációjával kapjuk, de a továbblépéshez azt sosem használjuk fel. Az aktív változatnál minden időpillanatban kombináljuk a két különböző lépésközű megoldást, de ez esetben a következő lépésben felhasználjuk az így nyert értéket, azzal számolunk tovább. Ezeknek a számítási módját az alábbi ábrákon láthatjuk.

A passzív Richardson-extrapoláció számítási módja:

$$\begin{tikzcd}
 & z_1 \arrow[r] & z_2 \arrow[r] & \cdots \\
y_0 \arrow[ur] \arrow[dr] & & & \\
 & w_1 \arrow[r] \arrow[d] & w_2 \arrow[r] \arrow[d] & \cdots \\
 & y_1 & y_2 &
\end{tikzcd}$$

Az aktív Richardson-extrapoláció számítási módja:

$$\begin{tikzcd}
 & z_1 \arrow[dr] & & z_2 \arrow[dr] & & \\
y_0 \arrow[ur] \arrow[dr] & & y_1 \arrow[ur] \arrow[dr] & & y_2 & \cdots \\
 & w_1 \arrow[ur] & & w_2 \arrow[ur] & &
\end{tikzcd}$$

Melyik változatot érdemes használni?

Erre a kérdésre nincs egy meghatározott válasz, három különböző eset állhat elő, melyeket gondosan mérlegelni kell és néhány alapvető tulajdonság vizsgálatával, az eredmények összevetésével eldönteni, hogy melyik változatot alkalmazzuk.

1. Mind az aktív, mind a passzív Richardson-extrapoláció alkalmazásával egy-egy új módszert kapunk (B módszer), amelyek ugyanolyan vagy legalább hasonló stabilitási tulajdonságokkal rendelkeznek, mint az eredeti numerikus módszer (A módszer).

2. Az új numerikus módszer (B módszer), amelyet a passzív Richardson-extrapoláció alkalmazásával nyerünk, jobb stabilitási tulajdonságokkal rendelkezik, mint az aktív Richardson-extrapolációval kapott (B) módszer.

3. Az új numerikus módszer (B módszer), amelyet az aktív Richardson-extrapoláció alkalmazásával nyerünk, jobb stabilitási tulajdonságokkal rendelkezik, mint a passzív Richardson-extrapolációval nyert (B) módszer.

A Richardson-extrapoláció alkalmazásával kapott eredményeket mindig egy teljesen új numerikus módszerként kell kezelni, ezért szükséges minden tulajdonságát alaposan megvizsgálni, legfőképp a stabilitást.

Ezt a fejezetet a [3], [6], [7] jegyzetek segítségével dolgoztuk fel.

6. Összefoglalás

Számos olyan módszer létezik, amellyel a közönséges differenciálegyenletek kezdeti-érték-feladatát numerikusan tudjuk megoldani. A dolgozatban a legfontosabb egylépéses módszereket ismertettük. Megvizsgáltuk, hogyan viselkednek az így kapott közelítő megoldások egy ekvidisztáns felosztású rácsháló pontjaiban. Láttuk, hogy meglehetősen általános feltételek mellett

- az explicit Euler-módszer elsőrendben konzisztens, és elsőrendben konvergens,
- az implicit Euler-módszer elsőrendben konzisztens,
- a trapézmódszer másodrendben konzisztens.

Ezeket a módszereket egy általánosabb képlettel is felírtuk, ez a θ-módszer. Egyszerű példával szemléltettük, hogy ha az implicit és az explicit Euler-módszer alkalmazása során egyre kisebb lépésközt választunk, akkor egyre közelebb kerülünk a pontos megoldás grafikonjához.

Ezek után a stabilitással foglalkoztunk. Ahhoz, hogy egy matematikai modell jól működjön, elengedhetetlen, hogy többféle értelemben is stabil legyen a megoldása. Az ún. 0-stabilitás a konzisztencia mellett biztosítja a konvergenciát. Fontos emellett a rögzített rácson való stabilitás is. Láttuk, hogy az implicit Euler-módszer A-stabil, erősen A-stabil és L-stabil, míg az explicit Euler-módszerre ezek a tulajdonságok nem teljesülnek.

Megismerkedtünk a Richardson-extrapolációval, melynek segítségével növelhetjük egy numerikus módszer pontosságát. Levezettük, hogyan kaphatunk egy p-ed rendben konvergens módszerből $(p+1)$-ed rendű módszert, majd megnéztük, mi történik, ha még egyszer alkalmazzuk az extrapolációt: ebből már $(p+2)$-ed rendben konvergens módszert nyertünk.

Becslést adtunk a közelítés során fellépő hibára, melynek segítségével formulát kaptunk a lépésköz automatikus szabályozására. Végül röviden ismertettük a Richardson-extrapoláció két változatát.

7. Függelék

A dolgozat során felhasznált MATLAB programkódok

Az explicit Euler-módszer

```
function [t, y] = szdexpeuler(szddiffegy, t0, y0, h, N)
t=zeros(N+1,1);
y=zeros(N+1,1);
t(1)=t0;
y(1)=y0;
for i=1:N
    t(i+1)=t(i)+h;
    y(i+1)=y(i)+h*szddiffegy(t(i),y(i));
end
u=exp(-t)+t;
plot(t,u,'b',t,y,'r')
xlabel('t')
ylabel('y')
title('Explicit Euler-módszer \tau lépésközzel')
legend('Pontos megoldás', 'Numerikus megoldás')
```

Az implicit Euler-módszer

```
function [t, y] = szdimpeuler(szddiffegy, t0, y0, h, N)
t=zeros(N+1,1);
y=zeros(N+1,1);
t(1)=t0;
y(1)=y0;
for i=1:N
    t(i+1) = t(i) + h;
    g = @(x)(x-y(i)-h*szddiffegy(t(i+1),x));
    g_der = @(x)(1+h);
```

```
    hiba = 1e-6;
    maxit = 50;
    x(1) = y(i)+h*szddiffegy(t(i),y(i));
    y_newton = g(x(1));
    it = 0;
    while ((abs(y_newton) > hiba) && (it < maxit))
        it = it + 1;
        x(it+1) = x(it) - g(x(it))/g_der(x(it));
        y_newton = g(x(it+1));
    end
    y(i+1)=x(it+1);
end
u=exp(-t)+t;
plot(t,u,'b',t,y,'r')
xlabel('t')
ylabel('y')
title('Implicit Euler-módszer \tau lépésközzel')
legend('Pontos megoldás','Numerikus megoldás')
```

A differenciálegyenletet megadó alfüggvény kódja

```
function dydt = szddiffegy(t,y)
dydt = -y+ t +1;
```

Hivatkozások

[1] I. Faragó: Alkalmazott analízis 2, előadás jegyzet, 2016 ősz

[2] P. Simon: Differenciálegyenletek, előadás jegyzet, 2016 tavasz

[3] Z. Zlatev, I. Dimov, I. Faragó, Á. Havasi: Richardson Extrapolation-Practical aspects and applications, De Gruyter Series in Applied and Numerical Mathematics 2 (2017)

[4] I. Faragó, R. Horváth: Numerikus módszerek (2013)

[5] I. Faragó: Numerikus modellezés és közönséges differenciálegyenletek numerikus megoldási módszerei (2013)

[6] B. Brajnovits: A Richardson-extrapoláció alkalmazása és matematikai vizsgálata egyszerűsített környezeti modellekben, szakdolgozat (2011) (http://nimbus.elte.hu/tanszek/docs/MSc/2011/BrajnovitsBrigitta_2011.pdf)

[7] Á. Havasi: A Richardson-extrapoláció elméleti vizsgálata és alkalmazásai (http://www.mettars.hu/wp-content/uploads/2013/05/Eloadas130530.pdf)

[8] G. Svantnerné Sebestyén: Alkalmazott analízis 2, gyakorlat jegyzet, 2016 ősz

[9] U. M. Ascher, L. R. Petzold: Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations (1997)

[10] L. F. Richardson: The Approximate Arithmetical Solution by Finite Differences of Physical Problems Including Differential Equations, with an Application to the Stresses in a masonry dam, Philosophical Transactions of the Royal Society of London, Series A (1911)

[11] L. F. Richardson: The Deferred Approach to the Limit, I. Single Lattice, Philosophical Transactions of the Royal Society of London, Series A (1927)