Numerieke methoden voor stelsels gewone differentiaalvergelijkingen

Prof. Dr. Marnix Van Daele

Deel II

Lineaire Meerstapsmethoden

Hoofdstuk 5

Lineaire-stabiliteitstheorie

5.1 Inleiding

In Voorbeeld 2.3.3 werd een convergente LMM toegepast op een testprobleem. Aan de hand van de numerieke resultaten stelden we vast dat er een zekere waarde h^* van de staplengte bestaat zo dat voor vaste $h > h^*$ de fout toenam als x toenam, terwijl voor vaste $h < h^*$ deze fout afnam. Tevens zagen we in Voorbeeld 2.3.4, waar echter geen LMM werd gebruikt, dat er methoden bestaan waarbij *alle* vaste positieve waarden van h, de fouten toenemen als x toeneemt. In zulke situaties accumuleren de fouten in een ongunstige wijze, m.a.w. we hebben te maken met een *stabiliteitsprobleem*. De enige vorm van stabiliteit die we tot nu hebben beschouwd is de nulstabiliteit, die controleerde hoe de fouten accumuleren in de limiet $h \to 0$. We hebben echter een stabiliteitstheorie nodig die toepasbaar is als h een vaste van nul verschillende waarde aanneemt.

5.2 Enkele voorbeelden

We beschouwen het scalaire testprobleem

$$y' = \lambda\, y\,, \quad x \in [0, b]\,, \quad y(0) = 1\,, \quad \lambda < 0\,. \tag{5.1}$$

De analytische oplossing is $y(x) = \exp(\lambda\, x)$. We zullen dit probleem numeriek oplossen m.b.v.

- de 2-staps Adams–Bashforth-methode

$$y_{n+2} = y_{n+1} + \frac{h}{2}\,(3\, f_{n+1} - f_n), \quad n = 0,\, 1,\, \ldots,\, N-2\,, \quad h = \frac{b}{N}\,. \tag{5.2}$$

- de 2-staps Nyström-methode

$$y_{n+2} = y_n + 2\, h\, f_{n+1}\,. \tag{5.3}$$

Beide methoden werden gebruikt voor de oplossing van (5.1) waarbij $\lambda = -1$ en $b = 2.2$ gekozen werden. Als startwaarden hebben we de exacte oplossingen in $x = 0$ en $x = h$ genomen : $y_0 = 1$ en $y_1 = \exp(-h)$. In Tabel 5.1 geven we de waarden van de GAF bij de verschillende punten voor $N = 11$.

x_n	Adams–Bashforth	Nyström
0	0	0
0.2	0	0
0.4	−279	−219
0.6	−424	−91.6
0.8	−512	−329
⋮	⋮	⋮
1.6	−528	−459
1.8	−494	127
2.0	−455	−555
2.2	−414	313

Tabel 5.1: $E_n = y(x_n) - y_n$ voor $n = 0, 1, \ldots, 11$ in eenheden 10^{-5}.

In Tabel 5.2 geven we de waarden voor de GAF maar nu bij $x_N = b = 2.2$ en voor verschillende waarden van N.

N	Adams–Bashforth	Nyström
11	−414	313
12	−347	−500
13	−296	185
14	−255	−338
⋮	⋮	⋮
37	−36.2	0.7
38	−34.3	−27
39	−32.6	−0.05
40	−30.9	24
⋮	⋮	⋮
66	−11.3	−7.2
67	−11.0	−1.8
68	−10.7	−6.7
69	−10.4	−1.7

Tabel 5.2: $E_N = y(x_N) - y_N$ voor $N = 11, 12, \ldots, 69$ in eenheden 10^{-5}.

De resultaten in Tabel 5.1 tonen aan dat voor beide methoden de afwijkingen min of meer uniform zijn bij de eerste verdelingspunten. Zij bewaren hetzelfde gedrag aan het einde van het integratiegebied voor de Adams-formule maar oscilleren in teken voor de Nyström-methode. Als we de resultaten in Tabel 5.2 analyseren, dan merken we bij de Nyström-methode twee verschillende gedragspatronen bij E_N : één voor even waarden van n en één voor oneven waarden. Voor beide groepen verkleinen de E_N met stijgende N, maar het oscillerende gedrag blijft. Om de redenen van het verschillend gedrag van deze fouten bij die twee methoden te achterhalen, zullen we de oplossingen meer in detail theoretisch bestuderen.

5.2.1 De Adams–Bashforth-methode

De vergelijking (5.1) levert de volgende differentievergelijking :

$$y_{n+2} - \left(1 + \frac{3}{2}\,H\right)\, y_{n+1} + \frac{1}{2}\,H\, y_n = 0\,, \quad h = \frac{b}{N}\,, \quad n = 0,\, 1,\, \ldots,\, N-2\,, \tag{5.4}$$

waarbij $H := \lambda\, h$. De algemene oplossing van die vergelijking luidt

$$y_n = C_1\, r_1^n + C_2\, r_2^n\,, \tag{5.5}$$

waarbij r_1 en r_2 de wortels zijn van de vierkantsvergelijking

$$r^2 - \left(1 + \frac{3}{2}\,H\right)\, r + \frac{1}{2}\,H = 0\,.$$

We bekomen

$$r_1 \;=\; \frac{1}{2} + \frac{3}{4}\,H + \sqrt{\frac{1}{4} + \frac{1}{4}\,H + \frac{9}{16}\,H^2}\,, \tag{5.6}$$

$$r_2 \;=\; \frac{1}{2} + \frac{3}{4}\,H - \sqrt{\frac{1}{4} + \frac{1}{4}\,H + \frac{9}{16}\,H^2} \quad = \frac{H}{2\,r_1}\,. \tag{5.7}$$

Merk ook op dat

$$\lim_{h\to 0} r_1 = 1 \quad \text{en} \quad \lim_{h\to 0} r_2 = 0\,,$$

wat precies de wortels zijn van de karakteristieke vergelijking $\rho(\xi) = \xi^2 - \xi$ behorend bij de methode (5.2). De constanten C_1 en C_2 in (5.5) volgen uit

$$\begin{cases} y_0 = 1 = C_1 + C_2\,, \\ y_1 = C_1\, r_1 + C_2\, r_2\,, \end{cases}$$

of

$$\begin{cases} C_1 = 1 - C_2\,, \\ C_2 = \dfrac{y_1 - r_1}{r_2 - r_1}\,. \end{cases} \tag{5.8}$$

Als we aannemen dat h voldoende klein is zodat Hß0 en de uitdrukking (5.6) voor r_1 in een reeks ontwikkelen, verkrijgen we

$$r_1 = 1 + H + \frac{1}{2}\,H^2 - \frac{1}{4}\,H^3 + \mathcal{O}(H^4)\,. \tag{5.9}$$

De reeksontwikkeling van $\exp(H)$ geeft anderzijds

$$\exp(H) = 1 + H + \frac{1}{2}H^2 + \frac{1}{6}H^3 + \mathcal{O}(H^4)\,.$$

Uit bovenstaande betrekkingen volgt dat

$$r_1 = \exp(H) - \frac{5}{12}\,H^3 + \mathcal{O}(H^4)\,.$$

M.b.v. de Newton-formule

$$(a+b)^n = a^n + n\,a^{n-1}\,b + \frac{n\,(n-1)}{2}\,a^{n-2}\,b^2 + \ldots$$

vinden we, door $a = \exp(H)$ en $b = -\frac{5}{12}\,H^3 + \mathcal{O}(H^4)$ te stellen, dat

$$r_1^n = \exp(n\,H) + n\,\left[-\frac{5}{12}\,H^3 + \mathcal{O}(H^4)\right]\,\exp((n-1)\,H)\,. \tag{5.10}$$

Daarenboven is

$$\exp((n-1)\,H) = \exp(n\,H)\exp(-H) = \exp(\lambda\,x_n)\,(1 + \mathcal{O}(H))$$

zodat

$$r_1^n = \exp(\lambda x_n)\,\left(1 - \frac{5}{12}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^4)\right)\,. \tag{5.11}$$

We kunnen een analoge omvorming realizeren voor r_2^n, maar omdat $r_2 = \mathcal{O}(H)$ vinden we dat $r_2^n = \mathcal{O}(H^n)$ en daardoor kan de tweede term in (5.5) voor grote n en Hß0 verwaarloosd worden. Zo bekomen we

$$\begin{aligned} y_n &= (1 - C_2)\,r_1^n + \mathcal{O}(H^n) \\ &= \exp(\lambda\,x_n)\,\left(1 - \frac{5}{12}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^4)\right)\,(1 - C_2)\,. \end{aligned} \tag{5.12}$$

Substitueren we $y_1 = \exp(\lambda\,h)$, r_1 en r_2 in (5.8), dan vinden we

$$C_2 = -\frac{5}{12}H^3 + \mathcal{O}(H^4)\,,$$

zodat uiteindelijk

$$y_n = \exp(\lambda\,x_n)\,\left(1 - \frac{5}{12}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^3)\right)\,. \tag{5.13}$$

De eerste term in de tweede factor staat voor de exacte oplossing, terwijl de volgende de geaccumuleerde fout weergeven; we kunnen dus schrijven dat

$$E_n = y(x_n) - y_n = \exp(\lambda\,x_n) - y_n = \exp(\lambda\,x_n)\,\left(\frac{5}{12}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^3)\right)\,. \tag{5.14}$$

Hieruit volgt dat E_n naar nul convergeert bij toenemende n voor constante h en vaste $\lambda < 0$. Dit gedrag vinden we terug in Tabel 5.1 (eerste kolom). Men kan gemakkelijk verifiëren dat bij verwaarlozing van de $\mathcal{O}(h^3)$ term in (5.14) de E_n-waarden corresponderend met de

laatste vier opgesomde waarden in Tabel 5.1 in eenheden 10^{-5} de waarden $-538, -496, -451$ en -406 krijgen, wat goed overeenstemt met de experimentele waarden in de tabel.

Wanneer we de waarde van x_n vasthouden, bvb. $x_n = x_N = b$ stellen, en het totaal aantal intervallen laten variëren, dan kan de geaccumuleerde globale fout E_N in $x_N = b$ geschreven worden, rekening houdend met $N = b/h$:

$$E_N = \exp(\lambda\, b)(\frac{5}{12}\,\lambda^3\, b^3\, N^{-2} + \mathcal{O}(N^{-3}))\ . \tag{5.15}$$

Dit verklaart de afname van $|E_N|$ voor stijgende N, zoals ook merkbaar in Tabel 5.2 (eerste kolom).

Uit de bovenstaande analyse kan de volgende conclusie getrokken worden. De resultaten bekomen met de Adams–Bashforth-methode convergeren naar de exacte oplossing; de hoofdreden hiervoor ligt in het feit dat de tweede wortel van de karakteristieke veelterm r_2 naar nul nadert als $h \to 0$. Dit laatste is een gevolg van het feit dat de tweede wortel van de eerste karakteristieke veelterm van de methode $\rho(\xi)$ nul is. Methoden die slechts één wortel op de eenheidscirkel hebben worden soms wel gezegd te voldoen aan *the strong root condition*.

5.2.2 De Nyström-methode

Wordt de Nyström-methode (5.3) toegepast op de testvergelijking, dan ontstaat

$$y_{n+2} - 2\,H\,y_{n+1} - y_n = 0\ ,$$

zodat r_1 en r_2 de wortels zijn van

$$r^2 - 2\,H\,r - 1 = 0\ ,$$

d.i.

$$r_1 = H + \sqrt{1+H^2}\,, \quad r_2 = H - \sqrt{1+H^2} = -\frac{1}{r_1}\ . \tag{5.16}$$

Ook hier is

$$\lim_{h\to 0} r_1 = 1, \quad \text{maar} \quad \lim_{h\to 0} r_2 = -1\,.$$

We vinden aldus de wortels van de eerste karakteristieke veelterm van de methode, vermits $\rho(r) = r^2 - 1$. Zo bekomt men

$$\begin{aligned} r_1 &= H + 1 + \frac{1}{2}\,H^2 - \frac{1}{8}\,H^4 + \ldots \\ &= 1 + H + \frac{1}{2}\,H^2 + \mathcal{O}(H^4) \\ &= \exp(H) - \frac{1}{6}H^3 + \mathcal{O}(H^4)\,. \end{aligned}$$

Hieruit volgt dat

$$\begin{aligned} r_1^n &= \exp(k\,H) + k\,\left(-\frac{1}{6}\,H^3 + \mathcal{O}(H^4)\right)\,\exp((k-1)\,H) \\ &= \exp(\lambda\,x_n)\,\left(1 - \frac{1}{6}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^4)\right)\,. \end{aligned} \tag{5.17}$$

Steunend op (5.16) vinden we dan ook

$$r_2^n = \left(-\frac{1}{r_1}\right)^n = (-1)^n\,\exp(-\lambda\,x_n)\,\left(1 + \frac{1}{6}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^4)\right)\,. \tag{5.18}$$

Er komt hier onmiddellijk een verschil t.o.v. de Adams–Bashforth-methode tot uiting, nl. r_2^n is niet verwaarloosbaar en dient expliciet in acht genomen te worden in de uitdrukking (5.5) van y_n. Uit (5.8) is afleidbaar dat

$$C_2 = -\frac{1}{12}\,H^3 + \mathcal{O}(H^4)\,,$$

en uit (5.5) vindt men

$$y_n = \exp(\lambda\,x_n)\,\left(1 - \frac{1}{6}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^3)\right) - \frac{1}{12}\,(-1)^n\,\exp(-\lambda\,x_n)\,(H^3 + \mathcal{O}(h^4))\,, \tag{5.19}$$

zodat

$$E_n = \exp(\lambda\,x_n)\,\left(\frac{1}{6}\,\lambda^3\,x_n\,h^2 + \mathcal{O}(h^3)\right) + \frac{1}{12}\,(-1)^n\,\exp(-\lambda\,x_n)\,\left(H^3 + \mathcal{O}(h^4)\right)\,. \tag{5.20}$$

Deze GAF bezit twee componenten, één evenredig met de exacte oplossing $\exp(\lambda\,x_n)$, een ander evenredig met de bijkomende functie $\exp(-\lambda\,x_n)$. Om de praktische gevolgen van die tweeledige structuur in te zien, beschouwen we opnieuw de resultaten uit de Tabellen 5.1 en 5.2.

We houden h constant en laten n stijgen. Het gedrag van de twee componenten wordt beheerst door de exponentiële functies $\exp(\pm\lambda\,x_n)$. Vermits λ negatief is, verkleint de eerste component met stijgende n, terwijl de tweede component, die alterneert in teken voor opeenvolgende waarden van n, toeneemt in absolute waarde. Die laatste component domineert meer en meer bij grotere n, wat het oscillerende karakter verklaart van de GAF in Tabel 5.1 (tweede kolom).

We kunnen ook kijken wat gebeurt in het punt $x_N = b$ als N varieert. We weten dat $h = b/N$ en herschrijven (5.20) als:

$$E_n = \exp(\lambda\,b)\,\left(\frac{1}{6}\,\lambda^3\,b^2\,N^{-2} + \mathcal{O}(N^{-3})\right) + \frac{1}{12}\,(-1)^N\,\exp(-\lambda\,b)(\lambda^3\,b^3\,N^{-3} + \mathcal{O}(N^{-4}))\,.$$

Deze formule verklaart het gedrag van de GAF in Tabel 5.2 (tweede kolom). Voor kleine waarden van n domineert de tweede component, zodat bij het begin de E_N alternerende tekens hebben bij opeenvolgende waarden van N. Als N echter toeneemt, dempt de tweede componente uit als N^{-3}, d.i. sneller dan de eerste component, zodat een kritische waarde N_c bestaat waarbij de twee componenten elkaar compenseren. Voor waarden van $N > N_c$

veranderen de E_N-waarden niet langer van teken, alhoewel ze wel sprongen vertonen bij opeenvolgende waarden van n omwille van het feit dat E_N de som is van de twee componenten voor even N en het verschil voor oneven N.

Het optreden van de tweede component volgt uit het feit dat r_2 nooit verwaarloosbaar is, wat op zijn beurt volgt uit het feit dat $\rho(\xi)$ naast de $+1$ wortel ook een wortel -1 bezit; zulke methoden worden soms gezegd te voldoen aan *the weak root condition.*

Wat we hier voor twee voorbeelden hebben vastgesteld zullen we in de volgende subparagraaf voor algemene LMM bestuderen.

5.3 De lineaire-stabiliteitstheorie

In verband met de gebruikte nomenclatuur wensen we er wel op te wijzen dat sommige auteurs i.p.v. *lineaire-* de benaming *zwakkestabiliteitstheorie* hanteren. In deze theorie maken we net als in vorige subparagraaf gebruik van een eenvoudig teststelsel, waarvoor al de oplossingen naar nul convergeren als x naar oneindig nadert. We zullen dan pogen voorwaarden te vinden opdat de numerieke oplossingen zich op een analoge wijze zouden gedragen. Het eenvoudigste teststelsel is het lineaire homogene stelsel met constante coëfficiënten:

$$y' = A\,y\,, \tag{5.21}$$

waarbij de eigenwaarden, $\lambda_t, t = 1,\, 2,\, \ldots,\, m$ van de constante $m \times m$ matrix (die we allemaal verschillend veronderstellen) voldoen aan

$$\mathrm{Re}\,\lambda_t < 0, \quad t = 1,\, 2,\, \ldots,\, m\,. \tag{5.22}$$

De algemene oplossing van (5.21) neemt dan de vorm aan:

$$y(x) = \sum_{t=1}^{m} \kappa_t\, \exp(\lambda_t\, x)\, c_t\,, \tag{5.23}$$

en uit (5.22) volgt dat alle oplossingen $y(x)$ van (5.21) voldoen aan

$$\|y(x)\| \to 0 \quad \text{als } x \to \infty\,.$$

De vraag die hier aan de orde is, luidt: *Welke voorwaarden moeten we opleggen opdat, wanneer een LMM toegepast wordt op (5.21), de numerieke oplossingen $\{y_n\}$ voldoen aan*

$$\|y_n\| \to 0 \text{ als } n \to \infty\,. \tag{5.24}$$

Als de LMM

$$\sum_{j=0}^{k} \alpha_j\, y_{n+j} = h \sum_{j=0}^{k} \beta_j\, f_{n+j} \tag{5.25}$$

toegepast wordt op (5.21) bekomen we voor $\{y_n\}$ het differentiestelsel

$$\sum_{j=0}^{k} (\alpha_j\, I_m - h\,\beta_j\, A)\, y_{n+j} = 0\,, \tag{5.26}$$

waarbij I_m de $m \times m$ eenheidsmatrix is. Vermits de eigenwaarden van A verschillend verondersteld zijn, bestaat er een niet-singuliere matrix Q zo dat

$$Q^{-1} A Q = \Lambda := \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_m] .$$

Definiëren we z_n als

$$y_n = Q z_n , \tag{5.27}$$

dan ontstaat door (5.26) links te vermenigvuldigen met Q^{-1}

$$\sum_{j=0}^{k} (\alpha_j I_m - h \beta_j \Lambda) z_{n+j} = 0 .$$

Vermits I_m en Λ beide diagonale matrices zijn, wordt dit een ongekoppeld stelsel en we kunnen schrijven dat

$$\sum_{j=0}^{k} (\alpha_j - h \beta_j \lambda_t) {}^t z_{n+j} = 0, \quad t = 1, 2, \ldots, m , \tag{5.28}$$

waarbij $z_n = [{}^1 z_n, {}^2 z_n, \ldots, {}^m z_n]^T$. Omdat de eigenwaarden van A in het algemeen complex zijn, stellen we vast dat elke vergelijking (5.28) een *complexe* lineair homogene differentievergelijking is met constante coëfficiënten. Uit (5.27) volgt dat $\|y_n\| \to 0$ als $n \to \infty$ impliceert dat $\|z_n\| \to 0$ als $n \to \infty$, wat betekent dat (5.24) voldaan is slechts en slechts dan als alle oplossingen $\{ {}^t z_n \}$ van (5.28) voldoen aan

$$| {}^t z_n| \to 0 \text{ als } n \to \infty, \quad t = 1, 2, \ldots, m . \tag{5.29}$$

Uit Stelling 1.6.1 weten we dat de algemene oplossing van elk van de differentievergelijkingen in (5.28) de volgende vorm aanneemt :

$${}^t z_n = \sum_{s=1}^{k} \kappa_{ts} r_s^n, \quad t = 1, 2, \ldots, m , \tag{5.30}$$

waarbij de κ_{ts} willekeurige complexe constanten zijn en r_s, $s = 1, 2, \ldots, k$ de verschillend veronderstelde wortels zijn van de karakteristieke veelterm

$$\sum_{j=0}^{k} (\alpha_j - h \beta_j \lambda_t) r^j .$$

Deze veelterm kan gemakkelijk geschreven worden in termen van de eerste en tweede karakteristieke veeltermen ρ en σ van de methode

$$\pi(r, \hat{h}) := \rho(r) - \hat{h} \sigma(r) , \quad \text{met } \hat{h} := h \lambda , \tag{5.31}$$

waarbij de complexe parameter λ een willekeurige eigenwaarde λ_t, $t = 1, 2, \ldots, m$ van A voorstelt. De veelterm $\pi(r, \hat{h})$ wordt de *stabiliteitsveelterm* van de methode genoemd. Het is duidelijk dat (5.29) en derhalve ook (5.24) voldaan zijn als alle wortels $r_s (= r_s(\hat{h}))$, $s = 1, 2, \ldots, k$ van $\pi(r, \hat{h})$ voldoen aan $|r_s| < 1$. Hierop steunend kunnen we de volgende definitie geven:

Definitie 5.3.1 *Een LMM wordt absoluut stabiel genoemd voor gegeven $\hat{h}$ als voor deze $\hat{h}$-waarde de wortels van de stabiliteitsveelterm voldoen aan $|r_s| < 1$, $s = 1, 2, \ldots, k$; zoniet wordt de LMM absoluut instabiel genoemd voor deze $\hat{h}$-waarde.* □

We wensen uiteraard te weten voor welke producten van h en λ de methode absoluut stabiel is. Vandaar de volgende definitie:

Definitie 5.3.2 *Een LMM bezit het gebied van absolute stabiliteit $\mathcal{R}_A$, waarbij $\mathcal{R}_A$ een gebied in het complexe $\hat{h}$-vlak, als ze absoluut stabiel is voor alle $\hat{h} \in \mathcal{R}_A$. De snijlijn van $\mathcal{R}_A$ met de reële as wordt het interval van absolute stabiliteit genoemd.* □

Merk op dat het interval van absolute stabiliteit relevant is voor het geval van de scalaire testvergelijking $y' = \lambda\, y$, met $\lambda \in \mathbb{R}$.

Het gebied van absolute stabiliteit, $\mathcal{R}_A$, is enkel een functie van de methode en de complexe parameter $\hat{h}$, zodat we in staat zijn voor elke LMM het gebied $\mathcal{R}_A$ in het complexe $\hat{h}$-vlak te tekenen. Als de eigenwaarden van de matrix A gekend zijn, is het mogelijk h voldoende klein te kiezen zo dat $h\,\lambda_t \in \mathcal{R}_A$ voor $t = 1, 2, \ldots, m$.

Uit (5.31) volgt dat $\pi(r, 0) = \rho(r)$. Bovendien is voor elke consistente LMM $\rho(1) = 0$ en $\rho'(1) \neq 0$. Dit betekent dat $\pi(1, 0) = 0$, en vermits de wortels van een veelterm continue functies zijn van de coëfficiënten van de veelterm, volgt hieruit dat er een wortel r_1 van π moet bestaan, die de eigenschap bezit dat $r_1 \to 1$ als $h \to 0$. Bovendien, vermits $\rho'(1) \neq 0$, is r_1 uniek.

Als de LMM convergent is, bezit de geassocieerde differentieoperator $\mathcal{L}$ orde $p \geq 1$. Dan is $\mathcal{L}[z(x);\, h] = \mathcal{O}(h^{p+1})$ voor elke voldoende afleidbare functie $z(x)$; $\exp(\lambda\, x)$ met $\lambda \in \mathbb{C}$ is zo een functie en we kunnen schrijven dat

$$\mathcal{L}[\exp(\lambda\, x);\, h] = \sum_{j=0}^{k} \{\alpha_j\, \exp[\lambda\,(x + j\,h)] - h\,\beta_j\,\lambda\, \exp[\lambda(x + j\,h)]\} = \mathcal{O}(h^{p+1})\,.$$

Na deling door $\exp(\lambda\, x)$ bekomen we

$$\sum_{j=0}^{k} \{\alpha_j\, [\exp(\hat{h})]^j - \hat{h}\,\beta_j\, [\exp(\hat{h})]^j\} = \mathcal{O}(\hat{h}^{p+1})\,,$$

wat ook kan geschreven worden als

$$\pi(\exp(\hat{h}),\, \hat{h}) = \mathcal{O}(\hat{h}^{p+1})\,. \tag{5.32}$$

Schrijven we nu $\pi(r,\, \hat{h})$ in termen van zijn wortels r_s, $s = 1, 2, \ldots, k$, dan bekomen we

$$\pi(r,\, \hat{h}) = (1 - \hat{h}\,\beta_k)\,(r - r_1)\,(r - r_2)\,\ldots\,(r - r_k)\,. \tag{5.33}$$

Opmerking 5.3.1

De factor $1-\hat{h}\,\beta_k$ nooit nul kan zijn: als $\tau(A)$ de spectraalstraal van A is, dan voldoet de Lipschitz-constante L van de functie $f = A\,y$ aan $L = \|A\| \geq \tau(A) \geq |\lambda|$, met λ een eigenwaarde van A. Uit Stelling A.4.1 volgt echter dat $h < 1/(|\beta_k|\,L)$, of m.a.w. dat $\hat{h} < 1/|\beta_k|$. □

Wanneer we nu in (5.33) $r = \exp(\hat{h})$ invullen en rekening houden met (5.32) bekomen we

$$[\exp(\hat{h}) - r_1]\,[\exp(\hat{h}) - r_2]\,\ldots\,[\exp(\hat{h}) - r_k] = \mathcal{O}(\hat{h}^{p+1})\;.$$

Vermits bij $\hat{h} \to 0$ geldt $\exp(\hat{h}) \to 1$ en $r_s \to \xi_s$ (met $\rho(\xi_s) = 0$), $s = 1,\,2,\,\ldots,\,k$, convergeert de eerste factor in het linkerlid naar nul als $h \to 0$ en geen enkele andere factor zal dit doen gezien $\rho'(1) \neq 0$.

Zo bekomen we

$$r_1 = \exp(\hat{h}) + \mathcal{O}(\hat{h}^{p+1})\;. \tag{5.34}$$

Voor waarden $\hat{h}$ in de omgeving van 0 met Re $\hat{h} > 0$ geldt dus steeds $|r_1| > 1$. Hieruit volgt onmiddellijk dat voor kleine $\hat{h}$ met Re $\hat{h} > 0$ een convergente LMM nooit absoluut stabiel kan zijn ! *Het gebied van absolute stabiliteit van elke convergente LMM kan de positieve reële as in de buurt van de oorsprong niet bevatten.*

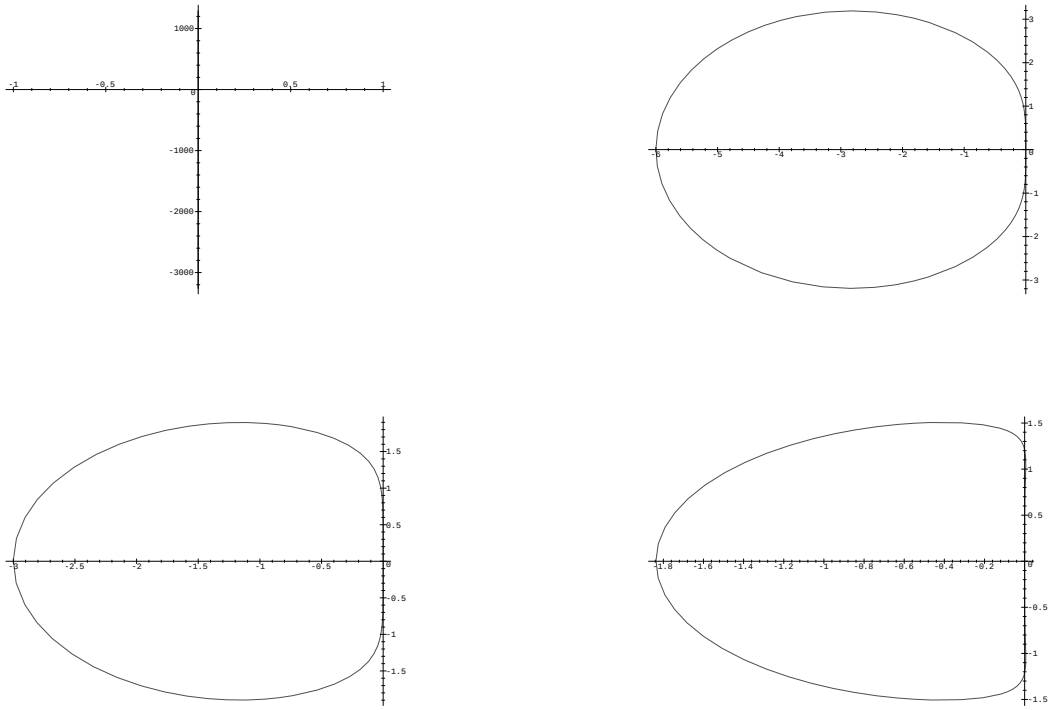
Merk wel op dat gezien bovenstaande argumenten asymptotisch zijn (d.i. voor $\hat{h} \to 0$) we niet kunnnen besluiten dat het gebied van absolute stabiliteit geen deel van de positieve reële as kan bevatten voor grote $|\hat{h}|$ of dat de grenzen van het gebied niet kunnen penetreren in het positieve halfvlak een eind weg van de oorsprong.

De meest handige manier om gebieden van absolute stabiliteit te vinden is de *boundary locus technique*. Het gebied $\mathcal{R}_A$ van het complexe $\hat{h}$–vlak wordt gedefinieerd door de voorwaarde dat voor alle $\hat{h} \in \mathcal{R}_A$ de wortels van $\pi(r,\,\hat{h})$ in modulus kleiner dan 1 zijn. We definiëren nu de omtrek $\partial\mathcal{R}_A$ in het complexe $\hat{h}$-vlak door te eisen dat voor alle $\hat{h} \in \partial\mathcal{R}_A$ één van de wortels r_i van $\pi(r,\,\hat{h})$ modulus 1 moet bezitten, m.a.w. van de vorm $r = \exp(i\,\theta)$ is. Vermits de wortels van een veelterm continue functies zijn van de coëfficiënten volgt hieruit dat de grenslijn van $\mathcal{R}_A$ moet bestaan uit $\partial\mathcal{R}_A$ (of een deel van $\partial\mathcal{R}_A$; sommige delen van $\partial\mathcal{R}_A$ zouden bijvoorbeeld kunnen correponderen met een $\pi(r,\,\hat{h})$ die één wortel met modulus 1 bezit, maar die daarenboven sommige wortels heeft met modulus kleiner dan één en sommige met modulus groter dan 1). Zo geldt voor alle $\hat{h} \in \partial\mathcal{R}_A$ de identiteit

$$\pi(\exp(i\,\theta),\,\hat{h}) = \rho(\exp(i\,\theta)) - \hat{h}\,\sigma(\exp(i\,\theta)) = 0\;. \tag{5.35}$$

Deze vergelijking is lineair in $\hat{h}$ en kan eenvoudig opgelost worden. De grenslijn $\partial\mathcal{R}_A$ wordt zo gegeven door

$$\hat{h} = \hat{h}(\theta) = \frac{\rho(\exp(i\,\theta))}{\sigma(\exp(i\,\theta))}\;. \tag{5.36}$$



Figuur 5.1: $\partial\mathcal{R}_A$ voor de k-staps Adams–Moulton-methoden met $k = 1$, 2, 3 en 4.

Voorbeeld 5.3.1

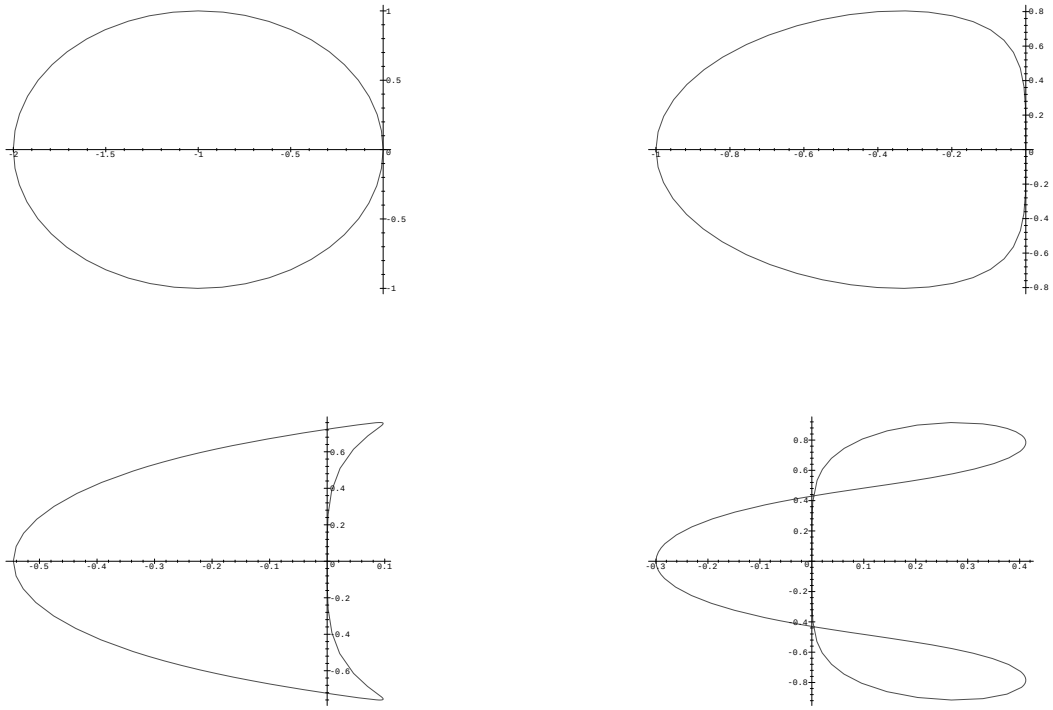
Voor de Adams–Moulton-methode met $k = 1$, de trapeziumregel, is

$$\hat{h}(\theta) = \frac{2\,i\,\sin\theta}{1+\cos\theta}$$

en $\mathcal{R}_A$ beslaat het ganse negatieve halfvlak. □

In de meeste gevallen zullen we (5.36) gebruiken om $\hat{h}(\theta)$ uit te tekenen voor een reeks $\theta \in [0,\, 2\,\pi]$ om daarna de uitgetekende punten te verbinden om een voorstelling van $\partial\mathcal{R}_A$ te verkrijgen. De randen $\partial\mathcal{R}_A$ die zo bekomen zijn voor de k-staps Adams–Moulton-methoden voor $k = 1, \ldots, 4$ zijn weergegeven in Figuur 5.1 en voor de Adams–Bashforth-methoden in Figuur 5.2.

Eenmaal de grenslijnen $\partial\mathcal{R}_A$ gevonden zijn, moet nog bepaald worden wat $\mathcal{R}_A$ is. Voor de Adams–Moulton-methode met $k = 2$, 3, 4 en voor de Adams–Bashforth-methode met $k = 1$, 2, 3 is $\partial\mathcal{R}_A$ een enkelvoudig gesloten grenslijn. Het vaststellen dat het inwendige van de gebieden begrensd door $\partial\mathcal{R}_A$ de gebieden van absolute stabiliteit zijn, volgt rechtstreeks uit het feit dat alle LMM absoluut instabiel zijn voor kleine positieve waarden van Re $\hat{h}$. Voor de Adams–Moulton-methode met $k = 1$ is de grenslijn $\partial\mathcal{R}_A$ niet gesloten; $\mathcal{R}_A$ is zoals reeds opgemerkt het volledige negatieve halfvlak. De analyse wordt veel ingewikkelder voor de Adams–Bashforth-methode met $k = 4$, waar de grenslijn $\partial\mathcal{R}_A$ wel gesloten is, maar



Figuur 5.2: $\partial\mathcal{R}_A$ voor de k-staps Adams–Bashforth-methoden met $k = 1$, 2, 3 en 4.

ingewikkeld qua vorm. Het bovengebruikte argument leidt er opnieuw toe dat $\mathcal{R}_A$ zeker reeds het begrensde deel in het linkerhalfvlak is, maar hiermee is het niet evident dat de andere twee begrensde stukjes, liggend in het rechterhalfvlak, moeten uitgesloten worden. Het punt op $\partial\mathcal{R}_A$ dat correspondeert met de waarde $\theta = \pi/2$ ligt in het eerste quadrant. Dit punt is $\hat{h} = 0.272 + 0.578\,i$. Met deze $\hat{h}$-waarde bezit $\pi(r, \hat{h})$ de wortels

$$i\,, \quad 1.076 + 0.744\,i\,, \quad 0.357 + 0.051\,i\,, \quad 0.190 - 0.470\,i\,.$$

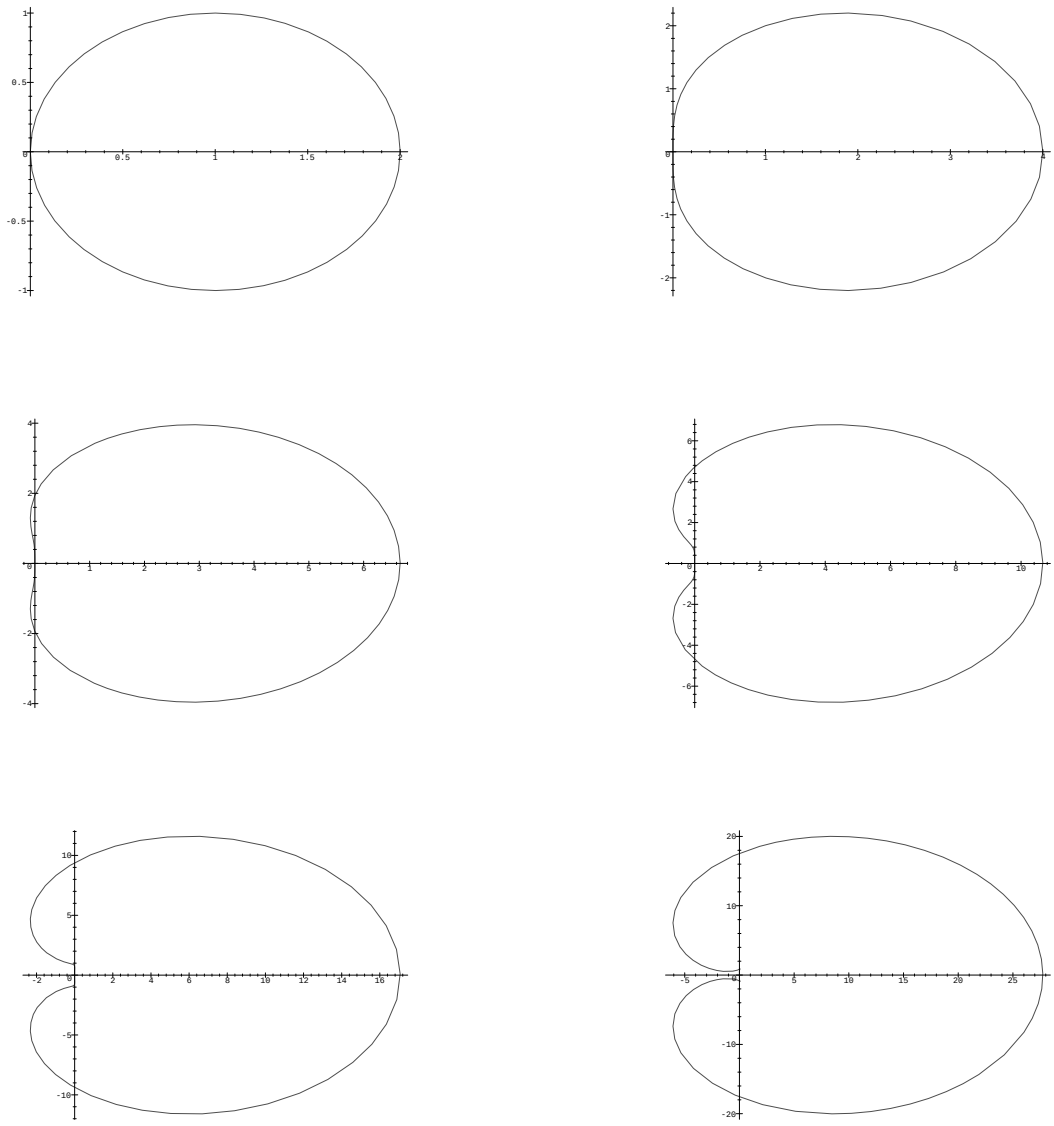
Merk wel op dat, vermits π een veelterm is met complexe coëfficiënten, de complexe wortels niet langer toegevoegde paren zijn. De eerste wortel bezit modulus 1 (dit is nodig want het punt behoort tot $\partial\mathcal{R}_A$), maar de tweede wortel is groter dan 1 in modulus. Door te steunen op de continuïteit kunnen we besluiten dat het inwendige van de curve geen gebied van absolute stabiliteit kan zijn. Door symmetrie kunnen we een zelfde conclusie trekken voor het gebied in het vierde kwadrant.

Uit Figuur 5.1 en Figuur 5.2 kunnen twee conjecturen geformuleerd worden :

1. *Impliciete methoden hebben grotere gebieden van absolute stabiliteit dan expliciete methoden.*

2. *De grootte van de gebieden van absolute stabiliteit verkleint als de orde toeneemt.*

De eerste conjectuur is waar voor alle numerieke methoden, de tweede echter niet. Voor RKM zal namelijk het omgekeerde blijken.



Figuur 5.3: $\partial\mathcal{R}_A$ voor de k-staps BDF-methoden met $k = 1$, 2, 3, 4, 5 en 6.

Het is in deze context dat het belang van de besproken BDF-methoden zich situeert. Deze methoden hebben zeer grote gebieden van absolute stabiliteit. Ze worden gegeven in Figuur 5.3 : $\mathcal{R}_A$ is telkens het onbegrensde deel. Noteer dat al deze gebieden voor $1 \leq k \leq 6$ de ganse negatieve reële as bevatten en dat voor $k = 1,\ 2$ ze zelfs het ganse negatieve halfvlak bevatten. Voor $k = 3$ dringt $\partial\mathcal{R}_A$ een heel klein beetje binnen in het linkerhalfvlak. Deze eigenschappen zijn belangrijk in de context van *stijve stelsels*.

Voorbeeld 5.3.2

Voor de Simpson-regel geldt

$$\rho(r) = r^2 - 1\,, \quad \sigma(r) = \frac{1}{3}\,(r^2 + 4\,r + 1)\,,$$

zodat

$$\hat{h}(\theta) = \frac{3\,i\,\sin\theta}{2 + \cos\theta}\,.$$

Hieruit volgt dat $\partial\mathcal{R}_A$ volledig op de imaginaire as ligt. Voor elke $\theta \in [0,\ 2\,\pi]$ is

$$-\sqrt{3} \leq \frac{3\sin\theta}{2 + \cos\theta} \leq \sqrt{3}\,.$$

Daarom is $\partial\mathcal{R}_A$ het deel van de imaginaire as gelegen tussen $-\sqrt{3}\,i$ en $\sqrt{3}\,i$. Het is gemakkelijk na te trekken dat $\pi(x,\ \hat{h})$ voor $\hat{h} \in \partial\mathcal{R}_A$ twee wortels met modulus 1 bezit. De Simpsonregel bezit dus een ledig gebied van absolute stabiliteit. Deze methode is een typisch voorbeeld voor een optimale methode; die optimale methoden bezitten gebieden van absolute stabiliteit die ofwel ledig ofwel onbruikbaar zijn omdat ze de negatieve reële as in de omgeving van de oorsprong niet bevatten. Dit verklaart waarom optimale methoden van weinig nut zijn. □

Het interval van absolute stabiliteit kan natuurlijk rechtstreeks afgeleid worden uit het gebied, maar soms wensen we enkel het interval te vinden, en daarvoor zijn snellere methoden beschikbaar. In dat geval is $\hat{h} \in \mathbb{R}$ waardoor $\pi(r, \hat{h})$ een reële veelterm wordt. In dit geval kan het *Routh–Hurwitz*-criterium gehanteerd worden.

Voorbeeld 5.3.3

Beschouw de drie-staps Adams–Moulton-methode gegeven door

$$\rho(r) = r^3 - r^2, \quad \sigma(r) = (9\,r^3 + 19\,r^2 - 5\,r + 1)/24\,.$$

Als we de notatie $H := \hat{h}/24$ invoeren, bekomen we

$$\pi(r, \hat{h}) = (1 - 9\,H)\,r^3 - (1 + 19\,H)\,r^2 + 5\,H\,r - H\,.$$

Door de transformatie $r = (1 + z)/(1 - z)$ bekomen we

$$(1 - z)^3\,\pi((1 + z)/(1 - z),\ \hat{h}) = a_0\,z^3 + a_1\,z^2 + a_2\,z + a_3\,,$$

waarbij

$$\begin{aligned} a_3 &= -24\,H > 0 \qquad \text{als } H < 0\,, \\ a_2 &= 2 - 48\,H > 0 \quad \text{als } H < \tfrac{1}{24}\,, \\ a_1 &= 4 - 16\,H > 0 \quad \text{als } H < \tfrac{1}{4}\,, \\ a_0 &= 2 + 16\,H > 0 \quad \text{als } H > -\tfrac{1}{8}\,. \end{aligned}$$

De voorwaarden $a_j > 0$, $j = 0, 1, 2, 3$ zijn voldaan a.s.a. $H \in]-1/8,\, 0[$. De overblijvende voorwaarde, $a_1\,a_2 - a_0\,a_3 > 0$ is voldaan a.s.a.

$$144\,H^2 - 22\,H + 1 > 0\,,$$

een voorwaarde die voor alle H steeds voldaan is. Hieruit volgt dat $\pi(r,\, \hat{h})$ Schur is a.s.a. $H \in]-1/8,\, 0[$ of $\hat{h} \in]-3,\, 0[$. Het interval van absolute stabiliteit is dus $]-3,\, 0[$, wat uiteraard ook afgelezen kan worden uit Figuur 5.1. □

5.4 Linearisatie van de foutvergelijking

De tot nu ingevoerde lineaire-stabiliteitstheorie is zeer restrictief, omdat we enkel het teststelsel $y' = A\,y$ beschouwen, terwijl we in de praktijk meestal geconfronteerd worden met algemene stelsels $y' = f(x,\, y)$. Er kan gepoogd worden de toepasbaarheid van de lineaire stabiliteitstheorie uit te breiden naar algemene stelsels door een aangepast benaderend stelsel voor differentievergelijkingen voor de GAF te construeren. Zo kan men uitgaan van de definitie van de LAF. Die impliceert dat

$$\sum_{j=0}^{k} \alpha_j\, y(x_n + j\,h) = h \sum_{j=0}^{k} \beta_j\, f(x_{n+j},\, y(x_{n+j})) + T_{n+k}\,.$$

De rij $\{y_n\}$ gegenereerd door de methode voldoet aan

$$\sum_{j=0}^{k} \alpha_j\, y_{n+j} = h \sum_{j=0}^{k} \beta_j\, f(x_{n+j},\, y_{n+j})\,.$$

Door die beide betrekkingen lid aan lid af te trekken en de Middelwaardestelling toe te passen, waarbij we $J = \partial f/\partial y$ stellen, verkrijgen we

$$\sum_{j=0}^{k} \alpha_j\, E_{n+j} = h \sum_{j=0}^{k} \beta_j\, \bar{J}(x_{n+j},\, \xi_{n+j})\, E_{n+j} + T_{n+k}\,. \tag{5.37}$$

De uitdrukking voor dit differentiestelsel voor $\{E_n\}$ is misleidend; dit stelsel lijkt lineair maar is het niet, vermits de waarden ξ_{n+j} allemaal liggen op het lijnsegment van y_{n+j} naar $y(x_{n+j})$ en daardoor is E_{n+j} een onbekende niet-lineaire functie van ξ_{n+j}. Zo'n stelsel kan niet opgelost worden; daarom forceren we het tot een lineair stelsel door de veronderstelling

$$\frac{\partial f}{\partial y} = J_C, \qquad \text{een constante matrix}\,. \tag{5.38}$$

Bovendien wordt verder verondersteld dat T_{n+k} een constante vector is : $T_{n+k} = T_C$. Hierdoor wordt (5.37)

$$\sum_{j=0}^{k} [\alpha_j \, I_m - h \, \beta_j \, J_C] \, E_{n+j} = T_C \; , \tag{5.39}$$

die ook wel de *gelinearizeerde foutvergelijking* wordt genoemd. Vermits de constante term T_C geen rol speelt in de bepaling of de normen van de oplossing van (5.39) stijgen of dalen in waarde als $n \to \infty$, kan deze verwaarloosd worden en (5.39) is essentieel hetzelfde stelsel als (5.26) waarin A vervangen is door J_C en y_{n+j} door E_{n+j}. De analyse gevoerd op de vorige pagina's blijft aldus geldig, zo dat we kunnen besluiten dat $\|E_n\| \to 0$ wanneer $n \to \infty$ als $h \, \lambda_t \in \mathcal{R}_A$, waarbij λ_t, $t = 1, 2, \ldots, m$ nu één van de eigenwaarden van J_C is en $\mathcal{R}_A$ het gebied van absolute stabiliteit van de methode. Deze bovenstaande redenering is echter niet altijd sluitend. De mogelijke fout in de bovenstaande redenering ligt in de veronderstelling (5.38). Het is niet waar dat de eigenwaarden van J_C altijd het gedrag van de oplossingen van het niet-lineaire stelsel (5.37) op een correcte wijze beschrijven. Enkel wanneer $f(x, \, y) = A \, y$, met A een constante matrix, is de redenering correct, vermits in dit geval de differentiestelsels voor de oplossing en voor de fouten essentieel dezelfde zijn.

5.5 Voorbeelden

We beschouwen vooreerst het probleem behandeld in de numerieke voorbeelden in paragraaf 2.3:

$$y' = f(x, \, y), \quad y(0) = \eta, \quad x \in [0, 1]$$

met

$$y = \begin{pmatrix} u \\ v \end{pmatrix} , \qquad f(x, \, y) = \begin{pmatrix} v \\ \frac{v \, (v-1)}{u} \end{pmatrix} , \qquad \eta = \begin{pmatrix} \frac{1}{2} \\ -3 \end{pmatrix} .$$

De unieke exacte oplossing is

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{8} \, (1 + 3 \, \exp(-8 \, x)) \\ -3 \, \exp(-8 \, x) \end{pmatrix} . \tag{5.40}$$

De Jacobiaan van dit stelsel is

$$J = \frac{\partial f}{\partial y} = \begin{bmatrix} 0 & 1 \\ -v \, (v-1)/u^2 & (2 \, v - 1)/u \end{bmatrix} .$$

De eigenwaarden van J zijn reëel :

$$\lambda_1 = \frac{v-1}{u} , \quad \lambda_2 = \frac{v}{u} . \tag{5.41}$$

Als we hierin de exacte oplossing voor u en v substitueren, bekomen we

$$\lambda_1 = -8, \quad \lambda_2 = -24/[3 + \exp(8 \, x)] . \tag{5.42}$$

In het integratieinterval [0, 1] varieert λ_2 van -6 naar ongeveer 0 zodat voor alle $x > 0$ beide eigenwaarden in het interval $[-8, 0]$ liggen. De LMM uit Voorbeeld 2.3.3 is

$$y_{n+3} + \frac{1}{4}\, y_{n+2} - \frac{1}{2}\, y_{n+1} - \frac{3}{4}\, y_n = \frac{h}{8}\,(19\, f_{n+2} + 5\, f_n)\ . \tag{5.43}$$

Vermits de eigenwaarden van de Jacobiaan reëel zijn, hoeven we enkel het interval (en niet het gebied) van absolute stabiliteit te bepalen. Dit interval is $]-1/3,\, 0[$ en derhalve kunnen we aan de voorwaarde

$$h\,\lambda_t \in \mathcal{R}_A, \quad t = 1,\, 2$$

voldoen door h zo te kiezen dat $-8\,h$ ligt in $]-1/3,\, 0[$. Dit kan door

$$h < h^* = \frac{1}{24} = 0.0417$$

te kiezen. Uit Tabel 2.3 zien we dat de GAF inderdaad afneemt voor $h < h^*$ en toeneemt voor $h > h^*$. De benadering met de gelinearizeerde foutvergelijking levert dus voor dit voorbeeld zinvolle resultaten.

We hoeven echter niet ver te zoeken om een tegenvoorbeeld te vinden : daartoe nemen we hetzelfde stelsel, maar met de beginwaarden

$$\eta = [-\frac{1}{4},\, 3]^T\ .$$

De exacte oplossing is nu

$$u(x) = [1 - 3\,\exp(-8\,x)]/8\,, \quad v(x) = 3\,\exp(-8\,x)\ . \tag{5.44}$$

Vermits het stelsel niet veranderd is, worden de eigenwaarden van de Jacobiaan nog altijd gegeven door (5.41), maar als we nu de exacte resultaten (5.44) voor u en v invoeren bekomen we

$$\lambda_1 = -8\,, \quad \lambda_2 = -24/[3 - \exp(8\,x)]\ .$$

Na vergelijking met (5.42) zien we dat λ_1 onveranderd gebleven is, maar het gedrag van λ_2 is radicaal veranderd voor positieve x-waarden. Bij $x = \hat{x} = (\ln 3)/8 \approx 0.137$ wordt λ_2 oneindig. Voor $x \in [0,\, \hat{x})$ wordt λ_2 negatief, maar $|\lambda_2|$ wordt zeer groot als $x \to \hat{x}$; voor $x > \hat{x}$ is λ_2 positief. Als we de numerieke oplossing van dit nieuwe IVP wensen te berekenen m.b.v. methode (5.43), waarvan het interval van absolute stabiliteit $]-1/3,\, 0[$ is, dan zou de theorie gebaseerd op de gelinearizeerde foutvergelijking voorspellen dat, om fouttoename te vermijden, we sterk afnemende waarden van h moeten voorzien als x nadert naar $\hat{x}$ en dat de fouttoename onvermijdelijk is als $x > \hat{x}$. In de praktijk zien we dat gedrag niet en de tabel van fouten bekomen voor numerieke oplossingen in het interval [0, 1] met een ruime keuze van staplengten is bijna identiek aan degene gegeven in Tabel 2.3 voor het originele IVP. De lineaire-stabiliteitstheorie voorspelt hier te negatieve resultaten.

De conclusie die we kunnen trekken uit dit voorbeeld mag niet zijn dat de lineaire-stabiliteitstheorie van geen waarde zou zijn. Een methode die de lineaire testvergelijking $y' = A\,y$ niet op een voldoening gevende wijze kan behandelen, is zeker geen goede kandidaat om in een code opgenomen te worden. De lineaire stabiliteitstheorie voorziet in een goed middel om verschillende klassen van methoden met elkaar te vergelijken.

Opmerking 5.5.1

Om de bovenstaande problemen op te heffen is er ook een niet-lineaire stabiliteitstheorie ontwikkeld. □

Opmerking 5.5.2

In codes wordt niet echt getest op absolute stabiliteit : naast het feit dat de bekomen resultaten niet altijd betrouwbaar zijn, is dit ook economisch onverantwoord vermits deze methode het voortdurend herberekenen van de Jacobiaan en zijn eigenwaarden vereist. In plaats daarvan betrouwen codes op het onder controle houden van de LAF. Wordt de schatting van de LAF te groot, dan wordt de stap geweigerd en de staplengte gereduceerd. □