Numerieke methoden voor stelsels gewone differentiaalvergelijkingen

Prof. Dr. Marnix Van Daele

Deel III

Runge–Kutta-methoden

Hoofdstuk 11

Lineaire-stabiliteitstheorie

11.1 Algemene inleiding

We stellen ons opnieuw de vraag : als $n \to \infty$, vertonen de componenten van de numerieke oplossing dan hetzelfde gedrag als de corresponderende componenten van de exacte oplossing? Neigen de numerieke componenten naar 0 als de exacte componenten dit doen?

In de lineaire-stabiliteitstheorie wordt aldus een teststelsel

$$y' = \mathcal{A}\, y \tag{11.1}$$

voorgesteld, waarbij $\mathcal{A}$ een $m \times m$ matrix is, met m verschillende eigenwaarden

$$\{\lambda_t \,|\mathrm{Re}\,(\lambda_t) < 0,\ t = 1,\, 2,\, \ldots,\, m\}\,.$$

Dit verzekert dat alle oplossingen van het teststelsel convergeren naar 0 als $x \to \infty$.

Vermits de eigenwaarden van $\mathcal{A}$ verschillend verondersteld worden, bestaat er een niet-singuliere matrix Q zodat

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

De transformatie $y = Q\,z$ leidt niet alleen tot een ontkoppeling van het originele stelsel (11.1). Net zoals bij LMM zorgt deze transformatie ook voor ontkoppeling van het differentiestelsel, voortvloeiend uit het gebruik van een RKM.

Wanneer we de algemene RKM (7.2) toepassen op (11.1) levert dit

$$\begin{cases} y_{n+1} = y_n + h \displaystyle\sum_{i=1}^{s} b_i\, k_i \\ \\ k_i = \mathcal{A}\left(y_n + h \displaystyle\sum_{j=1}^{s} a_{ij}\, k_j\right), \quad i = 1,\, 2,\, \ldots,\, s\,. \end{cases} \tag{11.2}$$

Definiëren we nu z_n en l_i door

$$y_n = Q\, z_n, \quad k_i = Q\, l_i, \quad i = 1,\, 2,\, \ldots,\, s$$

en substitueren we dit in (11.2) en vermenigvuldigen we achteraf links met Q^{-1}, dan bekomen we

$$\begin{cases} z_{n+1} = z_n + h \sum_{i=1}^{s} b_i\, l_i \\ l_i = \Lambda \left(z_n + h \sum_{j=1}^{s} a_{ij}\, l_j \right), \quad i = 1,\, 2,\, \ldots,\, s\,, \end{cases} \tag{11.3}$$

wat juist het resultaat zou zijn wanneer we de methode (7.2) zouden toepassen op het stelsel

$$z' = \Lambda\, z\,. \tag{11.4}$$

Hiermee is het duidelijk uit (11.3) en (11.4) dat we inderdaad het differentiaalstelsel en het differentiestelsel hebben ontkoppeld. Het is dus gerechtvaardigd als testvergelijking het scalaire probleem

$$y' = \lambda\, y, \quad \lambda \in \mathbb{C}, \quad \mathrm{Re}\ \lambda < 0 \tag{11.5}$$

te gebruiken. Als we de algemene RKM (7.2) toepassen op (11.5) zullen we steeds een 1-staps differentievergelijking bekomen van de vorm

$$y_{n+1} = R(\hat{h})\, y_n\,, \tag{11.6}$$

waarbij $\hat{h} = h\,\lambda$. We zullen $R(\hat{h})$ de *stabiliteitsfunctie* van de methode noemen. Als $n \to \infty$ zal $y_n \to 0$ a.s.a.

$$|R(\hat{h})| < 1 \tag{11.7}$$

en de methode is *absoluut stabiel* voor die waarden van $\hat{h}$ waarvoor (11.7) geldig is. Het gebied $\mathcal{R}_A$ van het complexe $\hat{h}$-vlak waarvoor (11.7) geldig is, is dan *het gebied van absolute stabiliteit* van de methode. De doorsnede van $\mathcal{R}_A$ met de reële as noemen we *het interval van absolute stabiliteit*.

Laat ons nu de vorm van $R(\hat{h})$ onderzoeken. Het is iets gemakkelijker te werken met de alternatieve vorm (7.5) van de algemene s-traps RKM.

Bij toepassing van de methode (7.5) op de testvergelijking (11.5), waarbij we weten dat y_n scalair is, vinden we

$$\begin{cases} Y_i = y_n + \hat{h} \sum_{j=1}^{s} a_{ij}\, Y_j, \quad i = 1,\, 2,\, \ldots,\, s \quad , \\ y_{n+1} = y_n + \hat{h} \sum_{i=1}^{s} b_i\, Y_i\,. \end{cases} \tag{11.8}$$

Zoals bij de bespreking van de subroutine PIRK in paragraaf 10.9 introduceren we Y en $e \in \mathbb{R}^s$ met $Y := [Y_1, Y_2, \ldots, Y_s]^T$ en $e := [1, 1, \ldots, 1]^T$.

Hierdoor kan (11.8) herschreven worden in de vorm

$$Y = y_n\, e + \hat{h}\, A\, Y, \quad y_{n+1} = y_n + \hat{h}\, b^T\, Y\ .$$

Uit de eerste vergelijking volgt dat

$$Y = y_n\, (I_s - \hat{h}\, A)^{-1}\, e$$

met I_s de $s \times s$ eenheidsmatrix. Substitutie in de tweede vergelijking levert

$$y_{n+1} = y_n\, [1 + \hat{h}\, b^T (I_s - \hat{h}\, A)^{-1}\, e]\ .$$

Hieruit volgt dat de stabiliteitsfunctie gegeven wordt door

$$R(\hat{h}) = 1 + \hat{h}\, b^T (I_s - \hat{h}\, A)^{-1} e\ . \tag{11.9}$$

Er bestaat een andere berekeningswijze voor de vorm van $R(\hat{h})$, aangebracht door Dekker en Verwer. We illustreren deze methode voor het geval $s = 2$. De betrekkingen (11.8) kunnen voor dit geval geschreven worden als

$$\begin{bmatrix} 1 - \hat{h}\, a_{11} & -\hat{h}\, a_{12} & 0 \\ -\hat{h}\, a_{21} & 1 - \hat{h}\, a_{22} & 0 \\ -\hat{h}\, b_1 & -\hat{h}\, b_2 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} y_n \\ y_n \\ y_n \end{bmatrix}\ .$$

Hieruit volgt als oplossing voor y_{n+1} m.b.v. de regel van Cramer

$$y_{n+1} = \frac{T}{N}$$

met

$$T = \det \begin{bmatrix} 1 - \hat{h}\, a_{11} & -\hat{h}\, a_{12} & y_n \\ -\hat{h}\, a_{21} & 1 - \hat{h}\, a_{22} & y_n \\ -\hat{h}\, b_1 & -\hat{h}\, b_2 & y_n \end{bmatrix}, \quad N = \det \begin{bmatrix} 1 - \hat{h}\, a_{11} & -\hat{h}\, a_{12} & 0 \\ -\hat{h}\, a_{21} & 1 - \hat{h}\, a_{22} & 0 \\ -\hat{h}\, b_1 & -\hat{h}\, b_2 & 1 \end{bmatrix}\ .$$

Als we in T de laatste rij aftrekken van elk van de voorgaande rijen, blijft de waarde van T onveranderd, d.w.z.

$$T = \det \begin{bmatrix} 1 - \hat{h}\, a_{11} + \hat{h}\, b_1 & -\hat{h}\, a_{12} + \hat{h}\, b_2 & 0 \\ -\hat{h}\, a_{21} + \hat{h}\, b_1 & 1 - \hat{h}\, a_{22} + \hat{h}\, b_2 & 0 \\ -\hat{h}\, b_1 & -\hat{h}\, b_2 & y_n \end{bmatrix} = y_n \det[I_s - \hat{h}\, A + \hat{h}\, e\, b^T]\ .$$

Rekening houdend met het feit dat $N = \det\, [I_s - \hat{h}\, A]$ en $y_{n+1} = R(\hat{h})\, y_n$ bekomen we dat

$$R(\hat{h}) = \frac{\det\, [I_s - \hat{h}\, A + \hat{h}\, e\, b^T]}{\det\, [I_s - \hat{h}\, A]}\ . \tag{11.10}$$

Het is evident dat de bovenstaande afleiding kan uitgebreid worden tot het geval van algemene s en dat (11.10) algemeen geldig is. De alternatieve vormen (11.9) en (11.10) zijn complementair en soms is de ene handig te gebruiken, soms is de andere het.

Laat ons onderzoeken welke vorm $R(\hat{h})$ aanneemt als de methode expliciet is. In dat geval is A strikt benedendriehoekig. In dat geval is ook de matrix $I_s - \hat{h}\,A$ benedendriehoekig maar al de elementen op de hoofddiagonaal zijn gelijk aan 1. Hieruit volgt dat det $[I_s - \hat{h}\,A] = 1$. Uit (11.10) volgt dan ook dat voor alle ERKMn de stabiliteitsfunctie een veelterm in $\hat{h}$ is. Voor IRKMn en DIRKMn is det $[I_s - \hat{h}\,A]$ niet langer 1, maar zelf een veelterm in $\hat{h}$, zodat de stabiliteitsfunctie een rationale functie in $\hat{h}$ wordt.

Als $R(\hat{h})$ een veelterm in $\hat{h}$ is, kan de voorwaarde (11.7) voor absolute stabiliteit nooit voldaan zijn als $|\hat{h}| \to \infty$. Hieruit volgt dat *alle ERKMn eindige gebieden van absolute stabiliteit bezitten*. Wanneer $R(\hat{h})$ een rationale functie is in $\hat{h}$ is het op zijn minst mogelijk dat (11.7) kan voldaan zijn wanneer $|\hat{h}| \to \infty$; dit houdt dus in dat *IRKMn en DIRKMn oneindige gebieden van absolute stabiliteit kunnen hebben*.

11.2 Expliciete methoden

Uit hoofdstuk 3 weten we dat wanneer bij expliciete methoden gestreefd wordt naar een maximum bereikbare orde voor een s-traps methode er een aantal vrije parameters beschikbaar blijven. Tot nu toe zijn die vrijheidsgraden nog niet gebruikt ten voordele van iets; de lineaire-stabiliteitstheorie zou hiervoor een interessant domein kunnen zijn. *Waarom zouden we die vrije parameters niet kiezen om het gebied van absolute stabiliteit te optimalizeren*? Laat ons dit proberen in het geval van de familie van 3-traps methoden van orde 3 die voldoen aan de rijsom-voorwaarde. De Butcher-matrix en de ordevoorwaarden zijn:

$$\begin{array}{c|ccc} c_1 & 0 & 0 & 0 \\ c_2 & c_2 & 0 & 0 \\ c_3 & c_3 - a_{32} & a_{32} & 0 \\ \hline & b_1 & b_2 & b_3 \end{array}$$

$$\begin{cases} b_1 + b_2 + b_3 = 1 \\ b_2\,c_2 + b_3\,c_3 = \dfrac{1}{2} \\ b_2\,c_2^2 + b_3\,c_3^2 = \dfrac{1}{3} \\ b_3\,a_{32}\,c_2 = \dfrac{1}{6} \end{cases} \tag{11.11}$$

Uit paragraaf 7.3 weten we dat er één 2-parameter en twee 1-parameter familie oplossingen voor (11.11) bestaan. De gemakkelijkste wijze om $R(\hat{h})$ te berekenen is gebruik te maken van (11.9) en d te definiëren als

$$d := (I_s - \hat{h}\,A)^{-1}\,e \quad \text{of} \quad (I_s - \hat{h}\,A)\,d := e\ ,$$

waaruit

$$\begin{bmatrix} 1 & 0 & 0 \\ -c_2\,\hat{h} & 1 & 0 \\ (a_{32}-c_3)\,\hat{h} & -a_{32}\,\hat{h} & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} .$$

Als oplossing hiervoor vinden we

$$\begin{cases} d_1 = 1 \\ -c_2\,\hat{h}\,d_1 + d_2 = 1 \\ (a_{32}-c_3)\,\hat{h}\,d_1 - a_{32}\,\hat{h}\,d_2 + d_3 = 1 \end{cases} \Longleftrightarrow \begin{cases} d_1 = 1 \\ d_2 = 1 + c_2\,\hat{h} \\ d_3 = 1 + c_3\,\hat{h} + c_2\,a_{32}\,\hat{h}^2 \end{cases}$$

Uit (11.9) volgt dan dat

$$\begin{aligned} R(\hat{h}) &= 1 + \hat{h}\,b^T d \\ &= 1 + (b_1+b_2+b_3)\,\hat{h} + (b_2\,c_2 + b_3\,c_3)\,\hat{h}^2 + b_3\,a_{32}\,c_2\,\hat{h}^3 . \end{aligned} \tag{11.12}$$

Door nu gebruik te maken van de ordevoorwaarden (11.11) vinden we dat

$$R(\hat{h}) = 1 + \hat{h} + \frac{\hat{h}^2}{2} + \frac{\hat{h}^3}{6}$$

voor *alle* 3-traps methoden van orde 3 die voldoen aan de rijsom-voorwaarde. We moeten er dus niet meer op hopen om bijzondere waarden voor de vrije parameters te kiezen en aldus de lineaire stabiliteitseigenschappen voor deze families van methoden te verbeteren.

Het bovenstaande resultaat kan als volgt veralgemeend worden. Als de s-traps ERKM orde p bezit, dan volgt uit paragraaf 8.4 onder de veronderstelling $y_n = y(x_n)$, dat de waarde y_{n+1}, gegeven door de methode wanneer ze wordt toegepast op de testvergelijking (11.5), verschilt van de Taylor-reeksontwikkeling van de exacte oplossing $y(x_{n+1})$ van (11.5) door een term van orde h^{p+1}. De ontwikkeling van de exacte oplossing volgt uit het herhaald afleiden van (11.5) en is

$$y(x_{n+1}) = y(x_n) + h\,\lambda\,y(x_n) + \frac{1}{2!}\,h^2\,\lambda^2\,y(x_n) + \ldots + \frac{1}{p!}\,h^p\,\lambda^p\,y(x_n) + \mathcal{O}(h^{p+1}) ,$$

waaruit dus volgt dat de numerieke oplossing moet voldoen aan

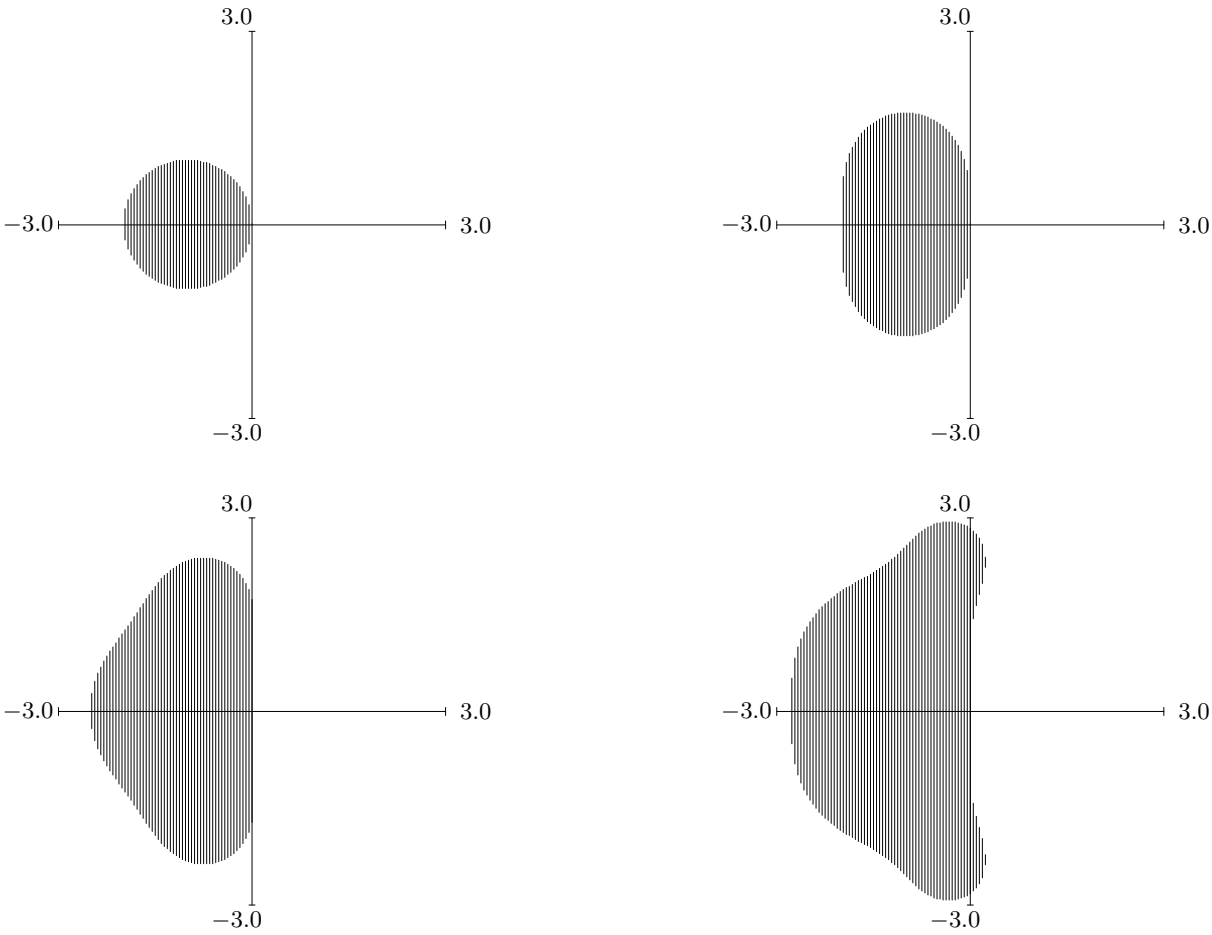
$$y_{n+1} = \left(1 + h\lambda + \frac{1}{2!}h^2\lambda^2 + \ldots + \frac{1}{p!}h^p\lambda^p\right) y_n + \mathcal{O}(h^{p+1})$$

of

$$y_{n+1}/y_n = 1 + \hat{h} + \frac{1}{2!}\,\hat{h}^2 + \ldots + \frac{1}{p!}\,\hat{h}^p + \mathcal{O}(\hat{h}^{p+1}) . \tag{11.13}$$

Anderzijds is het duidelijk uit (11.10) dat voor een s-traps methode $R(\hat{h})$ een veelterm is in $\hat{h}$ van graad ten hoogste s. Dit, samen met (11.13), impliceert dat als $s = p$ (en we weten dat dit zo is voor $1 \leq s \leq 4$) dat

$$R(\hat{h}) = y_{n+1}/y_n = 1 + \hat{h} + \frac{1}{2!}\,\hat{h}^2 + \ldots + \frac{1}{s!}\,\hat{h}^s . \tag{11.14}$$



Figuur 11.1: Stabiliteitsgebieden voor de s-traps ERK van orde s voor $s = 1, 2, 3, 4$.

We vinden aldus dat voor $1 \leq s \leq 4$ de stabiliteitsfuncties van alle s-traps ERKMn van orde s dezelfde structuur hebben. Ze worden gegeven door (11.14). In Figuur 11.1 worden de stabiliteitsgebieden van deze methoden getoond. Merk op dat voor $s = 1$ (de Euler-methode) de grenslijn van het gebied een cirkel is. Het is interessant te noteren dat bij stijgende orde het stabiliteitsgebied groter wordt. Merk op dat precies het tegengestelde gebeurt bij LMMn.

Opmerking 11.2.1

Een manier om die gebieden te bepalen kan gebeuren aan de hand van volgend algoritme. Als $\hat{h} = x + i\,y$, loop de lijn $x =$constante af en plaats een punt $(x,\, y)$ a.s.a. $|R(\hat{h})| < 1$; verhoog nadien x en herhaal het proces. Deze *scanning techniek* is alleen toepasbaar voor 1-staps methoden. □

Als de s-traps methode een orde $p < s$ bezit (en dit is altijd het geval als $s > 4$) dan volgt uit (11.13) en (11.14) dat de stabiliteitsfunctie de volgende vorm bezit:

$$R(\hat{h}) = 1 + \hat{h} + \frac{1}{2!}\,\hat{h}^2 + \ldots + \frac{1}{p!}\,\hat{h}^p + \sum_{q=p+1}^{s} \gamma_q\,\hat{h}^q\,, \tag{11.15}$$

waarbij de coëfficiënten γ_q functies zijn van de coëfficiënten van de methode. Eventueel kan hier een mogelijkheid bestaan om te pogen het stabiliteitsgebied te vergroten door een aangepaste invulling van de vrije parameters.

Opmerking 11.2.2

Beschouw een expliciete 3-trapsmethode van tweede orde. Hiervoor is

$$R(\hat{h}) = 1 + \hat{h} + \frac{1}{2!}\,\hat{h}^2 + \gamma_3\,\hat{h}^3\,.$$

Voor $\gamma_3 = 0$ bedraagt het stabiliteitsinterval $[-2,\,0]$, terwijl dit voor $\gamma_3 = 1/6$ (een noodzakelijke maar niet voldoende voorwaarde om een derde orde methode te verkrijgen) groeit tot $[-2.51,\,0]$. Voor $\gamma_3 = 1/12$ groeit dit interval zelfs tot $[-4.52,\,0]$. □

Echt spectaculaire verbeteringen heeft men op deze manier niet kunnen realiseren. We gaan daarom niet dieper in op deze problematiek. We kunnen echter wel nagaan hoe die γ_q er uitzien. De stabiliteitsfunctie kan als functie van $\hat{h}$ berekend worden uit (11.9). Als we opnieuw $(I_s - \hat{h}\,A)^{-1}\,e := d$ definiëren, dan levert

$$(I_s - \hat{h}\,A)\,d = e$$

een driehoekig stelsel dat gemakkelijk kan opgelost worden naar d. Bovendien weten we dat als de methode orde p bezit, $R(\hat{h}) = 1 + \hat{h}\,b^T\,d$ de vorm bezit gegeven door (11.15) waardoor we enkel de termen in $\hat{h}^q$ moeten vinden met $q = p,\,p+1,\,\ldots,\,s-1$ in d.

Laat ons dit illustreren door het bepalen van $R(\hat{h})$ voor een s-traps ERKM van orde $s-1$. In dit geval hoeven we enkel de term in $\hat{h}^{s-1}$ te vinden in d, wat betekent dat we enkel de hoogste graadsterm in $\hat{h}$ moeten kennen bij elke fase van de oplossing van het stelsel $(I_s - \hat{h}\,A)\,d = e$. Door de rijsom-voorwaarden in acht te nemen en de lagere-orde termen in $\hat{h}$ aan te geven met L.O. bekomen we :

$$\left\{\begin{array}{r} d_1 = 1 \\ -\hat{h}\,a_{21}\,d_1 + d_2 = 1 \\ -\hat{h}\,a_{31}\,d_1 - \hat{h}\,a_{32}\,d_2 + d_3 = 1 \\ \vdots \\ -\hat{h}\,a_{s1}\,d_1 - \ldots - \hat{h}\,a_{ss-1}\,d_{s-1} + d_s = 1 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} d_1 = 1 \\ d_2 = c_2\,\hat{h} + \text{L.O.} \\ d_3 = a_{32}\,c_2\,\hat{h}^2 + \text{L.O.} \\ \vdots \\ d_s = a_{s\,s-1}\,a_{s-1\,s-2}\,\ldots\,a_{32}c_2\,\hat{h}^{s-1} + \text{L.O.} \end{array}\right.$$

De term in $\hat{h}^s$ in $R(\hat{h}) = 1 + \hat{h}\,b^T\,d$ is dan $\gamma_s\hat{h}^s$ waarbij

$$\gamma_s = b_s\,a_{s\,s-1}\,a_{s-1\,s-2}\,\ldots\,a_{32}\,c_2\,,$$

wat in termen van de notatie in paragraaf 8.4 in Tabel 8.5, te schrijven valt als $\psi([_{s-1}\tau]_{s-1})$. We vinden aldus

$$R(\hat{h}) = 1 + \hat{h} + \frac{1}{2!}\,\hat{h}^2 + \ldots + \frac{1}{(s-1)!}\,\hat{h}^{s-1} + \psi([_{s-1}\tau]_{s-1})\,\hat{h}^s\,.$$

Merk op dat γ_s gemakkelijk berekend kan worden. Het is juist het product van de elementen op de eerste subdiagonaal van A vermenigvuldigd met b_s.

Let tevens op het feit dat als de methode orde s bezit, de ordevoorwaarden vereisen dat $\psi([_{s-1}\tau]_{s-1}) = 1/s!$, wat (11.14) bevestigt.

M.b.v. een gelijkaardige benadering (waar nu de twee hoogste machttermen in $\hat{h}$ beschouwd worden) kan men aantonen dat voor de s-traps methode van orde $p = s - 2$ geldt

$$R(\hat{h}) = 1 + \hat{h} + \frac{1}{2!}\,\hat{h}^2 + \ldots + \psi([_{s-2}\tau]_{s-2})\,\hat{h}^{s-1} + \psi([_{s-1}\tau]_{s-1})\,\hat{h}^s \ .$$

De uitbreiding voor het algemene geval $p < s$ is voor de hand liggend. M.b.v. deze benadering is het mogelijk de lineaire stabiliteitseigenschappen van de ingebedde methoden, besproken in paragraaf 9.3 te onderzoeken.

Voorbeeld 11.2.1

In wat volgt verwijzen het trapgetal s en de orde p naar de methode die de oplossing draagt en niet naar het paar ingebedde methoden.

(a) RKF45 : $s = 5$ en $p = 4$ met $\psi([_4\tau]_4) = \dfrac{1}{5}\dfrac{845}{4104}\dfrac{7296}{2197}\dfrac{9}{32}\dfrac{1}{4} = \dfrac{1}{104}$ waaruit

$$R(\hat{h}) = 1 + \hat{h} + \frac{\hat{h}^2}{2} + \frac{\hat{h}^3}{6} + \frac{\hat{h}^4}{24} + \frac{\hat{h}^5}{104}$$

(b) DOPRI (5,4) : $s = 6$ en $p = 5$ met $\psi([_5\tau]_5) = \dfrac{11}{84}\dfrac{5103}{18656}\dfrac{212}{729}\dfrac{32}{9}\dfrac{9}{40}\dfrac{1}{5} = \dfrac{1}{600}$ zodat

$$R(\hat{h}) = 1 + \hat{h} + \frac{\hat{h}^2}{2} + \frac{\hat{h}^3}{6} + \frac{\hat{h}^4}{24} + \frac{\hat{h}^5}{120} + \frac{\hat{h}^6}{600} \ .$$

□

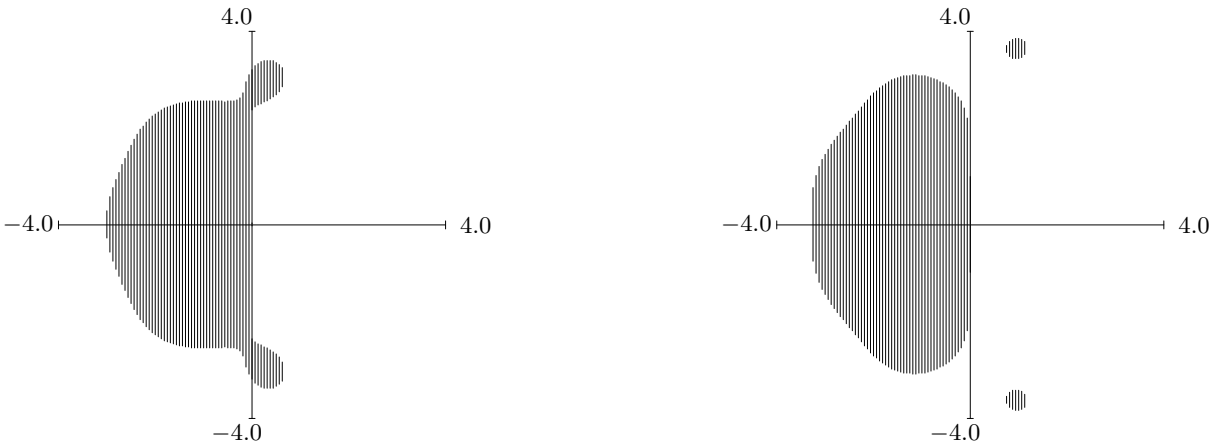
In de Figuur 11.2 worden de corresponderende gebieden van absolute stabiliteit weergegeven voor beide methoden. De aanwezigheid van een maan in de DOPRI(5,4) methode is merkwaardig.

11.3 Stijfheid en stabiliteitseigenschappen

Wanneer een methode in de lineaire-stabiliteitstheorie gekenmerkt wordt door een eindig stabiliteitsgebied, dan worden steevast voorwaarden opgelegd op $\hat{h} = \lambda\, h$. Deze voorwaarden drukken beperkingen uit op de stapgrootte : hoe groter $|\lambda|$, hoe kleiner $|h|$ moet gekozen worden om te voldoen aan de stabiliteitsvoorwaarde. Laat ons nu even veronderstellen dat we zo'n methode gebruiken om een lineair stelsel van de vorm

$$y' = \mathcal{A}\, y$$

op te lossen, waarbij minstens 1 eigenwaarde λ_0 van $\mathcal{A}$ gekenmerkt wordt door $\mathrm{Re}\,\lambda_0 \ll 0$. Dit kan er zelfs voor zorgen dat de stapgrootte h zo extreem klein gekozen moet worden om de stabiliteit te verzekeren, dat de aldus bekomen resultaten veel nauwkeuriger zijn dan



Figuur 11.2: Stabiliteitsgebieden voor RKF45 en DOPRI(5,4).

oorspronkelijk gevraagd. Heeft men te maken met een probleem waarbij dit het geval is, dan spreekt men van een *stijf probleem (stiff problem).*

Een exacte wiskundige omschrijving van het fenomeen stijfheid is heel moeilijk. Een diepgaande studie hiervan valt buiten het kader van deze cursus, maar de volgende omschrijvingen van het begrip stijfheid geven een idee van het probleem.

- Een lineair stelsel met constante coëfficiënten is stijf indien al haar eigenwaarden een negatief reëel deel hebben en indien de *stijfheidsverhouding (stiffness ratio)*

  $$\frac{\max_t |\mathrm{Re}\,\lambda_t|}{\min_t |\mathrm{Re}\,\lambda_t|}$$

  groot is.

- Stijfheid treedt op wanneer de stabiliteit en niet zozeer de nauwkeurigheid beperkingen oplegt aan de stapgrootte.

- Stijfheid treedt op wanneer sommige componenten van de oplossing sneller afnemen dan andere.

- Een stelsel wordt stijf genoemd in een bepaald interval indien in dat interval naburige oplossingscurves de oplossingscurve naderen aan een snelheid die zeer groot is in vergelijking met de snelheid waaraan de oplossing verandert aan dat interval.

Een poging om uit bovenstaande opmerkingen een valabele definitie te brouwen geeft volgend resultaat :

Definitie 11.3.1 *Indien een numerieke methode met een eindig gebied van absolute stabiliteit, toegepast op een stelsel met willekeurige beginwaarden, gedwongen wordt in een bepaald integratieinterval een stapgrootte te gebruiken die zeer klein is in vergelijking met de afleidbaarheid van de exacte oplossing in dat interval, dan wordt het stelsel in dat interval stijf genoemd.* □

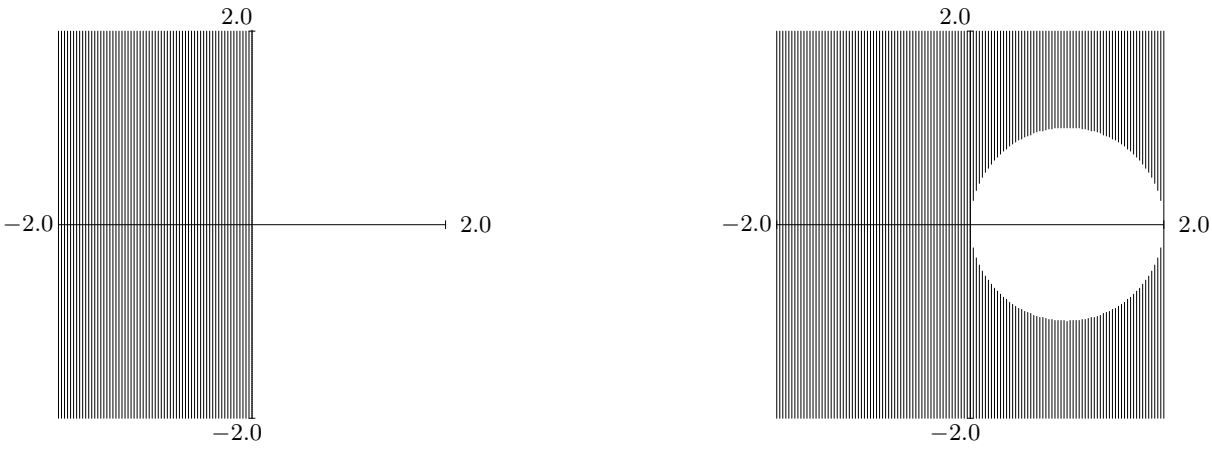
Gezien de strenge stabiliteitseisen die stijfheid oplegt aan de stapgrootte heeft men in de voorbije decennia heel wat aandacht besteed aan het stabiliteitsprobleem. Dit heeft geleid tot enkele belangrijke begrippen.

De exacte oplossing van het probleem

$$y' = \mathcal{A}\, y$$

is stabiel indien alle eigenwaarden van $\mathcal{A}$ behoren tot het linkse halfvlak $\mathbb{C}^-$. Het is dan ook een interessante eigenschap van numerieke methoden indien deze stabiliteitseigenschap behouden blijft. Dergelijke methoden werden door Dahlquist *A-stabiel* genoemd.

Definitie 11.3.2 *Een methode waarvoor het stabiliteitsgebied $\mathbb{C}^-$ bevat wordt A-stabiel genoemd.* □



Figuur 11.3: Stabiliteitsgebieden voor de trapeziumregel en de impliciete Euler-methode.

Voorbeeld 11.3.1

Een typisch voorbeeld voor een precies A-stabiele methode is de trapeziumregel

$$y_{n+1} = y_n + \frac{h}{2}\,(f_{n+1} + f_n)\,,$$

waarvoor de stabiliteitsfunctie gegeven wordt door

$$R(\hat{h}) = \frac{1 + \frac{\hat{h}}{2}}{1 - \frac{\hat{h}}{2}}\,,$$

waaruit

$$|R(x + i\,y)| = \left|\frac{(2 + x) + i\,y}{(2 - x) - i\,y}\right| < 1$$

a.s.a.

$$(2+x)^2 + y^2 < (2-x)^2 + y^2\,,$$

of na uitwerking $x < 0$.

Een andere A-stabiele methode is de impliciete Euler-methode

$$y_{n+1} = y_n + h\, f_{n+1}\,,$$

met stabiliteitsfunctie

$$R(\hat{h}) = \frac{1}{1-\hat{h}}\,.$$

De voorwaarde $|R(x+i\,y)| = 1$ definieert een cirkel met straal 1 en middelpunt $(1,0)$. Het stabiliteitsgebied is aldus meer dan het negatieve halfvlak alleen.

De stabiliteitsgebieden corresponderend met beide methoden worden getoond in Figuur 11.3 □

In de praktijk is gebleken dat A-stabiliteit (voor LMM) een zware voorwaarde betekent. Vandaar dat men ook gaan zoeken is naar teststelsels die minder strenge eisen opleggen, bvb. stelsels waarvan alle eigenwaarden λ_t voldoen aan $\pi - \alpha \leq \text{Arg } \lambda_t \leq \pi + \alpha$. Dit heeft geleid tot het begrip $A(\alpha)$-*stabiliteit*.

Definitie 11.3.3 *Een methode wordt $A(\alpha)$-stabiel genoemd met $\alpha \in (0, \pi/2)$ indien*

$$\{\hat{h} \mid -\alpha < \pi - \text{Arg } \hat{h} < \alpha\} \subseteq \mathcal{R}_A\,.$$

De methode wordt A(0)-stabiel genoemd indien de methode $A(\alpha)$-stabiel is voor een zekere $\alpha \in (0, \pi/2)$. □

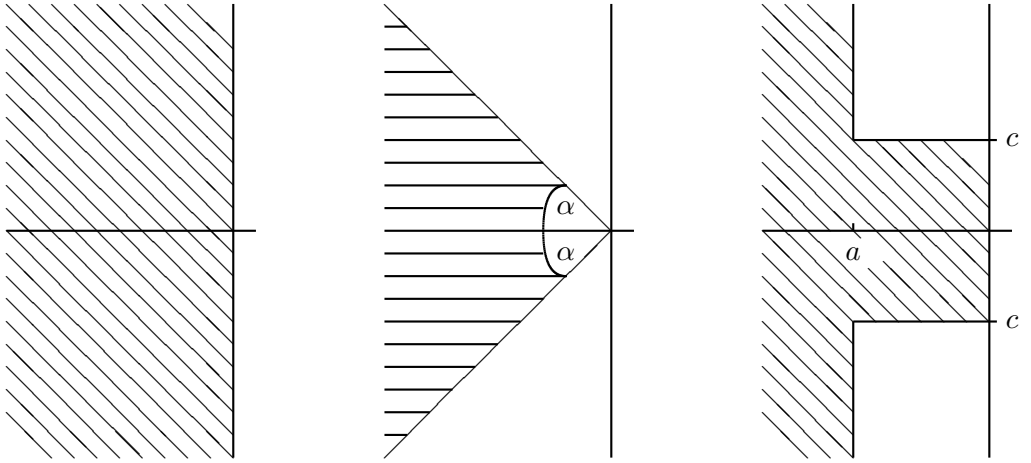
Het is duidelijk dat A(0)-stabiliteit een belangrijke eigenschap is voor problemen waarbij alle eigenwaarden reëel en negatief zijn. Voor deze problemen kunnen we echter ook eisen dat het stabiliteitsgebied de negatieve reële as bevat.

Definitie 11.3.4 *Een methode wordt A_0-stabiel genoemd indien*

$$\{\hat{h} \mid \text{Re } \hat{h} < 0,\ \text{Im } \hat{h} = 0\} \subseteq \mathcal{R}_A\,.$$

□

Tenslotte spreekt men ook soms van *stijve stabiliteit*. Stijfheid wordt voor vele problemen veroorzaakt door eigenwaarden met een groot negatief reëel deel, d.w.z. door eigenwaarden die links van de lijn $\text{Re } \hat{h} < -a$ met $a > 0$ liggen (en waarbij de overblijvende eigenwaarden dicht bij de oorsprong liggen). We nemen dus aan dat er geen eigenwaarden bestaan met een klein negatief reëel deel en een groot imaginair deel.



Figuur 11.4: A-stabiliteit, A(α)-stabiliteit en stijve stabiliteit

Definitie 11.3.5 *Een methode wordt stijf-stabiel genoemd indien* $\mathcal{R}_1 \cup \mathcal{R}_2 \subseteq \mathcal{R}_A$ *waarbij*

$$\mathcal{R}_1 = \{\hat{h} \,|\, \mathrm{Re}\, \hat{h} < -a\} \qquad \text{en} \qquad \mathcal{R}_2 = \{\hat{h} \,|\, -a < \mathrm{Re}\, \hat{h} < 0,\, -c \leq \mathrm{Im}\, \hat{h} \leq c\}$$

en waarbij a *en* c *positieve reële getallen zijn.* □

De minimale gebieden die nodig zijn om A-stabiliteit, A(α)-stabiliteit en stijve stabiliteit te verzekeren, worden getoond in Figuur 11.4.

Indien een methode een stabiliteitsgebied bezit dat precies samenvalt met $\mathbb{C}^-$, dan betekent dit dat $|R(i\,y)| = 1$ voor alle $y \in \mathbb{R}$. Vermits dan ook

$$\lim_{z \to -\infty} R(z) = \lim_{z \to \infty} R(z) = \lim_{z = i\,y,\, y \to \infty} R(z) = 1,$$

betekent dit dat voor z waarden met een zeer groot negatief reëel deel, $|R(z)|$ hoewel kleiner dan 1, dicht bij 1 ligt. Men kan aantonen dat dit tot gevolg heeft dat de stijve componenten van de exacte oplossing (d.w.z. deze corresponderend met eigenwaarden met een groot negatief deel, m.a.w. de componenten die snel vervallen) slechts heel langzaam vervallen in de numerieke oplossing.

Voorbeeld 11.3.2

We weten uit Voorbeeld 11.3.1 dat de trapeziumregel precies A-stabiel is. We passen deze methode toe op een teststelsel $y' = A\,y$ waarbij de m eigenwaarden van A verschillend zijn en een negatief reëel deel hebben en waarbij we $\bar{\lambda}$ de eigenwaarde noemen met het grootste reële deel (in absolute waarde). We bekomen aldus het stelsel differentievergelijkingen

$$y_{n+1} = B\,y_n\,, \quad B = (I_s - \frac{h}{2}\,A)^{-1}\,(I_s + \frac{h}{2}\,A)\,. \tag{11.16}$$

De algemene oplossing van (11.16) heeft de vorm

$$y_n = \sum_{t=1}^{m} K_t\,(\mu_t)^n\,d_t\,, \tag{11.17}$$

waarbij arbitraire K_t constanten zijn en waarbij μ_t en d_t de (verschillend onderstelde) eigenwaarden en eigenvectoren van B zijn. De numerieke oplossing y_n die opgeleverd wordt door (11.17) is een benadering voor de exacte oplossing

$$y(x_n) = \sum_{t=1}^{m} \kappa_t \exp(\lambda_t\, x_n)\, c_t = \sum_{t=1}^{m} \kappa_t \exp(\lambda_t\, x_0)[\exp(\lambda_t\, h)]^n\, c_t\,, \tag{11.18}$$

waarbij de c_i de eigenvectoren zijn van A. Heeft A de eigenwaarden λ_t dan weten we dat de eigenwaarden van $R(A)$ gegeven worden door $R(\lambda_t)$, m.a.w.

$$\mu_t = \frac{1 + \frac{h}{2}\lambda_t}{1 - \frac{h}{2}\lambda_t}\,, \qquad t = 1\,, 2\,, \ldots\,, m\,.$$

In het bijzonder moet B ook een eigenwaarde $\bar{\mu}$ bezitten gegeven door

$$\bar{\mu} = \frac{1 + \frac{h}{2}\bar{\lambda}}{1 - \frac{h}{2}\bar{\lambda}}\,. \tag{11.19}$$

Vergelijken we nu (11.17) met (11.18), dan zien we dat μ_t een benadering is voor $\exp(\lambda_t\, h)$. Vermits de methode A-stabiel is, is $|\mu_t| < 1$ zodat $(\mu_t)^n \to 0$ als $n \to \infty$. Maar, is $|\mathrm{Re}\,\bar{\lambda}|$ zeer groot en h niet buitengewoon klein (het idee achter A-stabiliteit is juist h niet te klein te moeten stellen), dan is $|h\,\bar{\lambda}|$ groot en $\bar{\mu} \approx -1$. We vinden dus dat de term $[\exp(\bar{\lambda}\, h)]^n$, die zeer snel uitdempt als $n \to \infty$, benaderd wordt door een term $(\bar{\mu})^n$, die slechts zeer traag uitdempt en alterneert van teken. We mogen dus een traag uitdempende oscillerende fout verwachten bij gebruik van de trapeziumregel in zulke omstandigheden.

Zouden we gebruik maken van de impliciete Euler-formule

$$y_{n+1} = y_n + h\, f_{n+1}\,,$$

dan komt dit er op neer dat B nu gegeven wordt door $B = (I_s - h\, A)^{-1}$, zodat

$$\bar{\mu} = \frac{1}{1 - h\,\bar{\lambda}}\,,$$

wat ongeveer 0 is indien $|h\,\bar{\lambda}|$ groot is. In dit geval zouden de termen $[\exp(h\,\bar{\lambda})]^n$ en $(\bar{\mu})^n$ beide snel naar 0 naderen voor $n \to \infty$ en we hoeven dan ook geen traag uitdempende fout te verwachten. □

Uit dit voorbeeld lijkt het aldus ook wenselijk te eisen dat $|R(z)|$ veel kleiner is dan 1 voor $z \to -\infty$.

Definitie 11.3.6 *(Ehle 1969) Een methode wordt L-stabiel (stijf-A-stabiel) genoemd indien de methode A-stabiel is en bovendien ook voldoet aan*

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

□

De impliciete Euler-regel is aldus een voorbeeld van een L-stabiele methode.

Uit de bovenstaande redenering zou men kunnen afleiden dat L-stabiele methoden te verkiezen zijn boven A-stabiele methoden, maar dit is niet altijd het geval. Er bestaan klasses van problemen (bvb. indien een eigenwaarde van het stelsel een positief reëel deel bezit) waar L-stabiele methoden leiden tot misleidende numerieke resultaten. Omdat het stabiliteitsgebied van een L-stabiele methode ook een deel van het rechtse halfvlak omvat, bestaan er waarden voor $\hat{h}$ waarvoor de numerieke oplossing stabiel blijft (d.w.z. naar 0 nadert voor $n \to \infty$), terwijl de exacte oplossing in norm groter wordt voor $n \to \infty$.

Samenvattend kunnen we stellen dat A-stabiele methoden die niet L-stabiel zijn voor de meeste problemen zullen leiden tot oplossingen met traag uitdempende fouten, maar deze fouten kunnen voldoende onder controle gehouden worden door automatische controle van de stapgrootte. L-stabiele methoden daarentegen produceren geen traag uitdempende fouten maar kunnen tot misleidende (niet door automatische controle van de stapgrootte te ontdekken) fouten leiden voor zeldzame klassen van problemen. We besluiten dan ook dat *precies A-stabiele* methoden zowat de voorkeur genieten : zij garanderen dat de numerieke oplossing van het testprobleem $y' = \lambda\, y$ naar 0 convergeert voor $n \to \infty$ a.s.a. de exacte oplossing naar 0 nadert voor x naderend naar ∞.

11.4 Padé-benaderingen en ordesterren

We weten dat, wanneer een RKM toegepast wordt op de testvergelijking voor absolute stabiliteit

$$y' = \lambda\, y, \qquad \lambda \in \mathbb{C},$$

de vergelijking

$$y_{n+1} = R(\hat{h})\, y_n \tag{11.20}$$

wordt bekomen, waarbij $\hat{h} := h\,\lambda$ en waarbij R een rationale functie is in het geval van IRKMn of DIRKMn en een polynoom voor ERKMn.

De exacte oplossing van de testvergelijking (11.5) luidt

$$y(x) = K\, \exp(\lambda\, x)$$

met K een arbitraire constante. Hieruit volgt dat

$$y(x_{n+1}) = \exp(h\,\lambda)\, y(x_n) = \exp(\hat{h})\, y(x_n)\ . \tag{11.21}$$

Als we nu een RKM gebruiken van orde p en rekening houden met de veronderstelling $y_n = y(x_n)$, dan volgt uit $y_{n+1} = R(\hat{h})\, y_n$ en (11.21) dat

$$y(x_{n+1}) - y_{n+1} = [\exp(\hat{h}) - R(\hat{h})]\, y(x_n) = \mathcal{O}(h^{p+1})\ ,$$

waaruit

$$R(\hat{h}) = \exp(\hat{h}) + \mathcal{O}(h^{p+1})\ . \tag{11.22}$$

Door dit resultaat worden we gemotiveerd om de rationale benadering te bestuderen van de exponentiële functie $\exp(q)$, $q \in \mathbb{C}$. We definiëren daartoe de functie $R_T^S(q)$ met $S \geq 0$ en $T \geq 0$ als

$$R_T^S(q) = \frac{\sum_{i=0}^{S} a_i\, q^i}{\sum_{j=0}^{T} b_j\, q^j}, \quad a_0 = b_0 = 1, \quad a_S \neq 0, \quad b_T \neq 0\,, \tag{11.23}$$

waarbij a_i, $b_j \in \mathbb{R}$, $i = 0, 1, \ldots, S$, $j = 0, 1, \ldots, T$. Men zegt dat $R_T^S(q)$ een (S, T) *rationale benadering is van orde* p *voor de exponentiële functie* $\exp(q)$ als

$$R_T^S(q) = \exp(q) + \mathcal{O}(q^{p+1})\,.$$

Uit (11.23) volgt dat de orde van een gegeven rationale benadering p is als

$$(1 + a_1\, q + \ldots + a_S\, q^S) - (1 + b_1\, q + \ldots + b_T\, q^T)\left(1 + q + \frac{q^2}{2!} + \ldots\right) = \mathcal{O}(q^{p+1})\,. \tag{11.24}$$

Voorbeeld 11.4.1

Als $S = 2$ en $T = 2$ volgt uit de gelijkstelling van de coëfficiënten van q^k voor $k = 0, 1, 2, 3, 4$ het volgende stelsel:

$$\begin{cases} q^0: & 1 - 1 = 0 \\ q^1: & a_1 - b_1 = 1 \\ q^2: & a_2 - b_2 - b_1 = \dfrac{1}{2} \\ q^3: & \dfrac{b_1}{2} + b_2 = -\dfrac{1}{6} \\ q^4: & \dfrac{b_1}{6} + \dfrac{b_2}{2} = -\dfrac{1}{24} \end{cases}$$

De oplossing hiervan luidt:

$$a_1 = \frac{1}{2}, \quad a_2 = \frac{1}{12}, \quad b_1 = -\frac{1}{2}, \quad b_2 = \frac{1}{12}\,,$$

en

$$R_2^2(q) = \frac{1 + \frac{1}{2}\, q + \frac{1}{12}\, q^2}{1 - \frac{1}{2}\, q + \frac{1}{12}\, q^2}\,.$$

De benadering is op die wijze van orde 4. □

1	$1+q$	$1+q+\frac{1}{2}q^2$	$1+q+\frac{1}{2}q^2+\frac{1}{6}q^3$
$\frac{1}{1-q}$	$\frac{1+\frac{1}{2}q}{1-\frac{1}{2}q}$	$\frac{1+\frac{2}{3}q+\frac{1}{6}q^2}{1-\frac{1}{3}q}$	$\frac{1+\frac{3}{4}q+\frac{1}{4}q^2+\frac{1}{24}q^3}{1-\frac{1}{4}q}$
$\frac{1}{1-q+\frac{1}{2}q^2}$	$\frac{1+\frac{1}{3}q}{1-\frac{2}{3}q+\frac{1}{6}q^2}$	$\frac{1+\frac{1}{2}q+\frac{1}{12}q^2}{1-\frac{1}{2}q+\frac{1}{12}q^2}$	$\frac{1+\frac{3}{5}q+\frac{3}{20}q^2+\frac{1}{60}q^3}{1-\frac{2}{5}q+\frac{1}{20}q^2}$
$\frac{1}{1-q+\frac{1}{2}q^2-\frac{1}{6}q^3}$	$\frac{1+\frac{1}{4}q}{1-\frac{3}{4}q+\frac{1}{4}q^2-\frac{1}{24}q^3}$	$\frac{1+\frac{2}{5}q+\frac{1}{20}q^2}{1-\frac{3}{5}q+\frac{3}{20}q^2-\frac{1}{60}q^3}$	$\frac{1+\frac{1}{2}q+\frac{1}{10}q^2+\frac{1}{120}q^3}{1-\frac{1}{2}q+\frac{1}{10}q^2-\frac{1}{120}q^3}$

Tabel 11.1: De Padé-tabel

Men kan verwachten uit de telling van het aantal parameters a_i, b_i in (11.24) dat de maximum orde die een $R_T^S(q)$ benadering kan bereiken $S+T$ is. Zulke benaderingen van maximum orde zijn bekend als *Padé-benaderingen* en we zullen ze noteren als $\hat{R}_T^S(q)$. In Tabel 11.1 geven we de eerste Padé-benaderingen voor $\exp(q)$ in de zgn. *Padé-tabel*.

De lineaire stabiliteitseigenschappen van een 1-stapsmethode die de differentievergelijking (11.20) genereert, worden bepaald door het gedrag van $R(\hat{h})$. Daartoe introduceerde Ehle de benaming *aanvaardbaarheid (acceptibility)* :

Definitie 11.4.1 *Een rationale benadering $R(q)$ van $\exp(q)$ wordt*

- *A-aanvaardbaar genoemd indien $|R(q)| < 1$ zodra $\mathrm{Re}\, q < 0$.*
- *A_0-aanvaardbaar genoemd indien $|R(q)| < 1$ zodra $q \in \mathbb{R}$ en $q < 0$.*
- *L-aanvaardbaar genoemd indien $R(q)$ A-aanvaardbaar en bovendien $|R(q)| \to 0$ voor $\mathrm{Re}\, q \to -\infty$.*

□

Het is duidelijk dat een methode aldus A-stabiel, A_0-stabiel of L-stabiel is a.s.a. $R(\hat{h})$ A-aanvaardbaar, A_0-aanvaardbaar of L-aanvaardbaar is. Het is duidelijk dat een rationale approximatie $R_T^S(q)$ niet A-aanvaardbaar kan zijn als $S > T$ en dat als $R_T^S(q)$ A-aanvaardbaar is en $T > S$, $R_T^S(q)$ dan ook L-aanvaardbaar is.

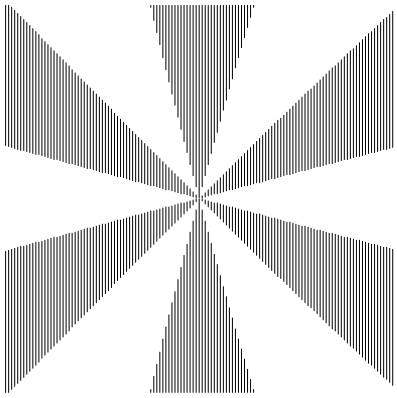
In verband met de Padé-benaderingen $\hat{R}_T^S(q)$ werden de volgende resultaten bewezen :

Stelling 11.4.1 *(Birkhoff en Varga, 1965) $\hat{R}_T^T(q)$ is A-aanvaardbaar.* □

Stelling 11.4.2 *(Varga, 1966) Als $T \geq S$, dan is $\hat{R}_T^S(q)$ A_0-aanvaardbaar.* □

Stelling 11.4.3 *(Ehle, 1969) Als $T = S+1$ of $T = S+2$, dan is $\hat{R}_T^S(q)$ L-aanvaardbaar.* □

Alle elementen op en onder de hoofddiagonaal van de Padé-tabel zijn aldus A_0-aanvaardbaar. De elementen op de hoofddiagonaal en de twee nevendiagonalen eronder zijn A-aanvaardbaar. Alle A-aanvaardbare elementen onder de hoofddiagonaal zijn automatisch ook L-aanvaardbaar. Tevens weten we dat er boven de hoofddiagonaal geen A-aanvaardbare elementen te vinden zijn. We blijven dus voorlopig nog geconfronteerd met de vraag of er nog A-aanvaardbare elementen zijn onder de tweede nevendiagonaal. De *conjectuur van Ehle* stelde dat dit niet zo was, zodat $\hat{R}_T^S(q)$ pas A-aanvaardbaar is voor $T-2 \leq S \leq T$. Deze conjectuur is vele jaren onopgelost gebleven. Uiteindelijk werd een zeer elegant bewijs geleverd door Wanner, Hairer en N rsett. Zij steunden daartoe op hun theorie van de *ordesterren*.



Figuur 11.5: Een ordesterren in de omgeving van de oorsprong. $\mathcal{B}$ bestaat uit de niet-gearceerde gebieden.

Het gebied van absolute stabiliteit $\mathcal{R}_A$ van een RK-methode of van een lineaire 1-staps methode wordt bepaald door

$$\mathcal{R}_A = \{q \in \mathbb{C} \,|\, |R(q)| < 1\}\,.$$

Wanner, Hairer en N rsett bestudeerden echter het gebied

$$\mathcal{B} = \{q \in \mathbb{C} \,|\, |R(q)| > |\exp(q)|\}\,.$$

Tot hun grote verbazing vonden ze bij het uitplotten van dit gebied rond 0 steeds een stervormige figuur zoals in Figuur 11.5 terug waarvan de vorm te maken had met de orde van de methode. Dit zette hen ertoe aan deze figuren, bepaald door $\mathcal{B}$ en zijn complement $\mathrm{C}(\mathcal{B})$, *ordesterren* te noemen.

Lemma 1 *$R(q)$ is een rationale benadering van $\exp(q)$ van orde p a.s.a. voor $q \to 0$, $\mathcal{B}$ bestaat uit $p+1$ sectoren gescheiden door $p+1$ sectoren van $C(\mathcal{B})$ en waarbij de hoek van elke sector $\pi/(p+1)$ is.* □

Bewijs. Zij

$$\frac{R(q)}{\exp(q)} = 1 - C\,q^{p+1} + \mathcal{O}(q^{p+2})\,.$$

Dan zal $R(q)/\exp(q)$ met $q = r\,e^{i\theta}$ (waarbij r voldoende klein wordt verondersteld) evenveel keer rond 1 draaien als q^{p+1} rond de oorsprong, namelijk $p+1$ keer. Dit betekent dat $R(q)/\exp(q)$ dan $p+1$ keer alternerend binnen en buiten de eenheidscirkel ligt. ($R(q)/\exp(q)$ ligt binnen de cirkel voor kleine positieve q waarden als $C > 0$.) ∎

Lemma 2 *De grens $\partial\mathcal{B}$ van $\mathcal{B}$ bezit juist 2 takken die naar oneindig gaan. Indien*

$$R(q) = K\,q^l + \mathcal{O}(q^{l-1}) \text{ voor } q \to \infty$$

dan benaderen deze takken asymptotisch

$$x = \log|K| + l\,\log|y|\,. \tag{11.25}$$

□

Bewijs. Zij $q = r\,e^{i\theta}$ waarbij r voldoende groot wordt verondersteld. Voor Re $q \to +\infty$ stijgt de exponentiële functie $\exp(q)$ sneller dan om het even welke polynomiale of rationale functie $R(q)$. Dit betekent dat $q \notin \mathcal{B}$ voor $-\pi/2 < \theta < \pi/2$. Anderzijds daalt de exponentiële functie $\exp(q)$ voor Re $q \to -\infty$ sneller dan om het even welke polynomiale of rationale functie $R(q)$. Dit betekent dat $q \in \mathcal{B}$ voor $\pi/2 < \theta < 3\,\pi/2$.
Om de uniciteit van de grenslijnen te bepalen beschouwen we voor r voldoende groot de twee functies

$$\begin{aligned}\varphi_1(\theta) &= |\exp(q)|^2 = \exp(2\,r\,\cos\theta)\\ \varphi_2(\theta) &= |R(q)|^2 = R(r\,e^{i\,\theta})\,R(r\,e^{-i\,\theta})\,.\end{aligned}$$

Uit

$$\varphi_1'(\theta) = -2\,r\,\sin\theta\,\exp(2\,r\,\cos\theta) = -2\,r\,\sin\theta\,\varphi_1(\theta)$$

en

$$\begin{aligned}\varphi_2'(\theta) &= r\,i\,e^{i\,\theta}\,R'(r\,e^{i\,\theta})\,R(r\,e^{-i\,\theta}) - r\,i\,e^{-i\,\theta}\,R(r\,e^{i\,\theta})\,R'(r\,e^{-i\,\theta})\\ &= r\left(i\,e^{i\,\theta}\,\frac{R'(r\,e^{i\,\theta})}{R(r\,e^{i\,\theta})} + (-i)\,e^{-i\,\theta}\,\frac{R'(r\,e^{-i\,\theta})}{R(r\,e^{-i\,\theta})}\right)\,\varphi_2(\theta)\\ &= 2\,r\,\mathrm{Re}\left(i\,e^{i\,\theta}\,\frac{R'(r\,e^{i\,\theta})}{R(r\,e^{i\,\theta})}\right)\end{aligned}$$

volgt

$$\frac{\mathrm{d}}{\mathrm{d}\,\theta}\log\varphi_1(\theta) = -2\,r\,\sin\theta\,, \qquad \frac{\mathrm{d}}{\mathrm{d}\,\theta}\log\varphi_2(\theta) = 2\,r\,\mathrm{Re}\left(i\,e^{i\,\theta}\,\frac{R'(r\,e^{i\,\theta})}{R(r\,e^{i\,\theta})}\right)\,.$$

Omdat voor voldoend grote r de verhouding R'/R tot nul nadert , geldt voor $\theta \in [\epsilon,\,\pi-\epsilon]$:

$$\frac{\mathrm{d}}{\mathrm{d}\,\theta}\log\varphi_1(\theta) < \frac{\mathrm{d}}{\mathrm{d}\,\theta}\log\varphi_2(\theta)$$

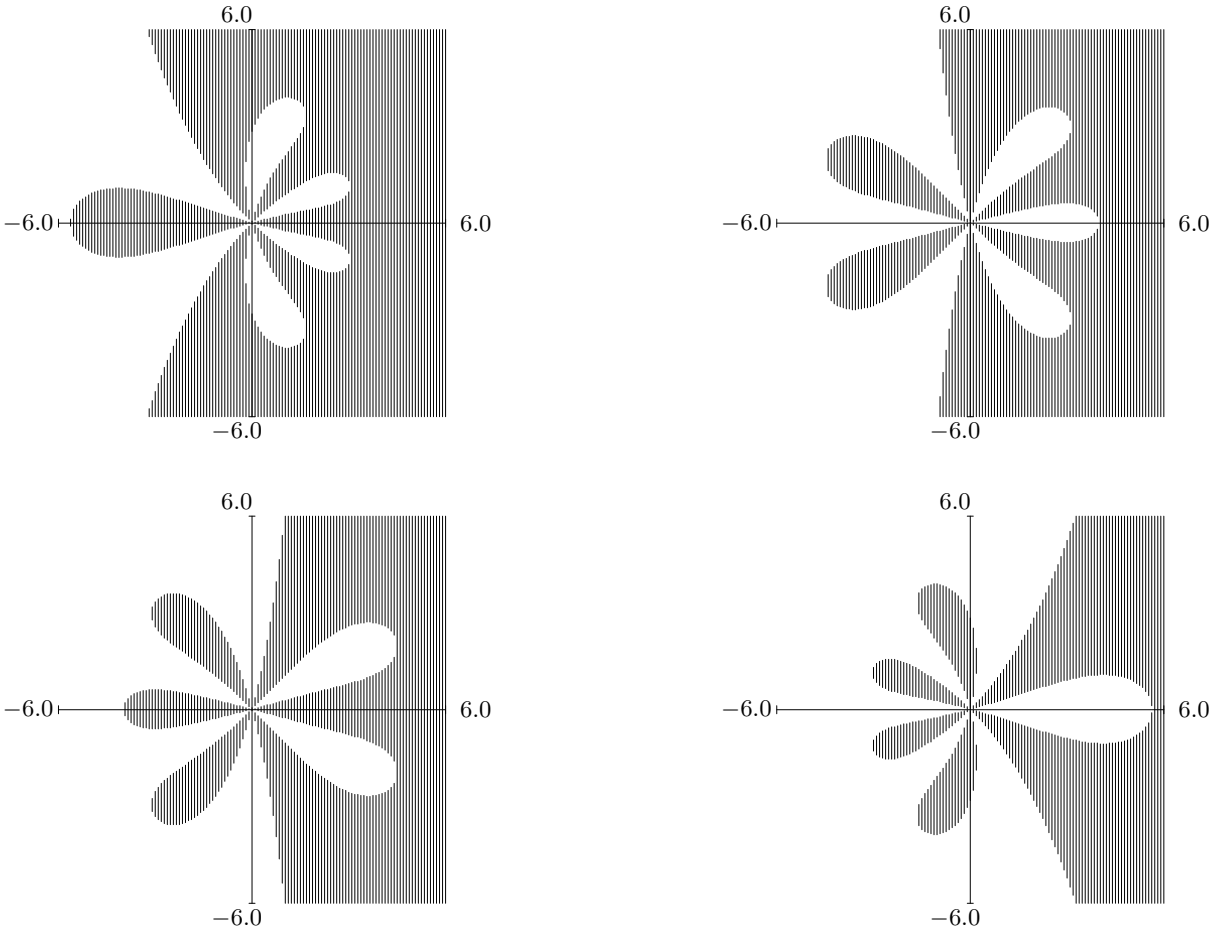
Bijgevolg is er slechts 1 waarde van θ mogelijk waarvoor $\varphi_1(\theta) = \varphi_2(\theta)$. Dit toont meteen de uniciteit aan van de twee takken.
De formule (11.25) volgt uit

$$|K|\,(x^2+y^2)^{l/2} \approx |e^q| = e^x\,,$$

zodat

$$\log|K| + \frac{l}{2}\,\log(x^2+y^2) \approx x\,.$$

Houden we nog rekening met het feit dat langs $\partial\mathcal{B}$ $x/y \to 0$ voor $x + i\,y \to \infty$, dan vinden we (11.25). ∎

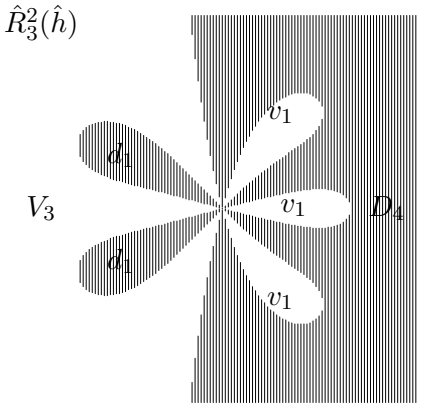


Figuur 11.6: Ordesterren voor $\hat{R}_4^1(\hat{h})$, $\hat{R}_3^2(\hat{h})$, $\hat{R}_2^3(\hat{h})$ en $\hat{R}_1^4(\hat{h})$. $\mathcal{B}$ bestaat uit de niet-gearceerde gebieden.

De beide eigenschappen die in voorgaande Lemma 1 en Lemma 2 zijn geformuleerd, worden geïllustreerd in Figuur 11.6, die de ordesterren voor Padé-benaderingen van orde 5 voorstellen voor grotere gebieden rond 0.

Gebieden van $\mathcal{B}$ die afkomstig zijn van één enkele sector worden vingers genoemd. We zullen zo'n vinger het etiket v_1 geven indien hij begrensd is en als V_1 noteren in het andere geval.

Analoog is de duale vinger d_1 een begrensde sector van $\mathrm{C}(\mathcal{B})$ en de duale vinger D_1 een onbegrensde sector van $\mathrm{C}(\mathcal{B})$. Indien een gebied n sectoren bevat van $\mathcal{B}$ dan wordt dit gebied een vinger van multipliciteit n genoemd en genoteerd als v_n of V_n. Analoog definiëren we ook d_n en D_n. Deze notaties worden geïllustreerd in Figuur 11.7. Merk op dat deze werkwijze er toe leidt dat de som van de indices van de letters v, V, d en D samen $2\,(p+1)$ bedraagt.



Figuur 11.7: Aanduiding van de vingers en de duale vingers van $\hat{R}_3^2(\hat{h})$.

Lemma 3 *Elke begrensde vinger van multipliciteit n bevat minstens n polen van $R(q)$ (multipliciteiten meegeteld), en elke begrensde duale vinger van multipliciteit n bevat minstens n nulpunten van $R(q)$ (multipliciteiten meegeteld).* □

Bewijs. Veronderstel dat de grens van een vinger v voorgesteld kan worden door de geparametriseerde positief georiënteerde curve $c(t)$ voor $t_0 \leq t \leq t_1$, waarbij $\vec{a} = (c_1(t),\, c_2(t))$ de raakvector is en $\vec{n} = (c_2(t),\, -c_1(t))$ de uitwendige normaalvector. Noteren we voor $q = x + i\,y$

$$f(q) = \frac{R(q)}{\exp(q)} = r(x,\, y)\, e^{i\,\varphi(x,\,y)}\,,$$

dan is $\log f(q) = \log r(x,\, y) + i\,\varphi(x,\, y)$. Vermits $\vec{n}$ een uitwendige normaal vector is t.o.v. de vinger v, is

$$0 > \frac{\partial\,(\log r)}{\partial\,\vec{n}} = \frac{\partial\,(\log r)}{\partial\,x}\,c_2(t) - \frac{\partial\,(\log r)}{\partial\,y}\,c_1(t)\,.$$

De Cauchy-Riemann vergelijkingen voor $\log f$ zijn nu echter

$$\frac{\partial\,(\log r)}{\partial\,x} = \frac{\partial\,\varphi}{\partial\,y} \qquad \text{en} \qquad \frac{\partial\,(\log r)}{\partial\,y} = -\frac{\partial\,\varphi}{\partial\,x}$$

waardoor

$$\frac{\partial\,\varphi}{\partial\,\vec{a}} < 0\,,$$

m.a.w. het argument van f neemt af langs de curve c. Stel nu dat de contourcurve $c(t)$ n keer terugkeert naar de oorsprong (waar het argument een veelvoud is van $2\,\pi$), dan maakt de vector $q(z)$ minstens n volledige omwentelingen in de negatieve zin. Cauchy's argument principe (dat zegt dat het aantal nulpunten (resp. polen) van de functie $f(q)$ in een gesloten contour gelijk is aan het aantal positieve (resp. negatieve) rotaties van de vector (Re $f(q)$, Im $f(q)$) als q langs de contour beweegt in positieve zin) leert dan dat er minstens n polen liggen in v (er zijn geen nulpunten want die kunnen niet in $\mathcal{B}$ liggen).
Als de grenscurve voorgesteld wordt door verschillende curves, dan worden de rotatiegetallen gewoon opgeteld. Voor duale vingers is het bewijs analoog, maar nu is $\partial\,(\log r)/\partial\,\vec{n} > 0$. ∎

Lemma 4 *Een niet-constante functie $R(q)$ is A-aanvaardbaar a.s.a. $\mathcal{B}$ niet snijdt met de imaginaire as en indien $R(q)$ geen polen bezit in $\{z \in \mathbb{C} \mid Re\, z < 0\}$.* □

Bewijs. Stel dat $R(q)$ A-aanvaardbaar is. Dan geldt voor iedere $q \in \mathbb{C}^-$ dat $|R(q)| < 1$. Derhalve kan $R(q)$ geen polen bezitten in dit gebied. Gezien langs de imaginaire as $|\exp(q)| = 1$ is, geldt voor deze punten meteen ook $|R(q)/\exp(q)| = |R(q)| \leq 1$, m.a.w. deze punten kunnen niet tot $\mathcal{B}$ behoren.
Voor de omgekeerde redenering maken we gebruik van het maximum modulus principe (dat zegt dat, als een functie in een open gebied holomorf is en een maximum (in norm) bezit, deze functie constant is in dat gebied). Gezien $R(q)$ een rationale functie is in q zonder polen in $\mathbb{C}^-$, is $R(q)$ holomorf in het open gebied $\mathbb{C}^-$. Anderzijds geldt, gezien $\mathcal{B}$ geen doorsnede bezit met de negatieve as, dat $|R(i\,q)| \leq 1$. Gezien het maximum $R(q)$ in $\mathbb{C}^-$ wegens het maximum modulus principe bereikt wordt langs de grens van $\mathbb{C}^-$, geldt dan $|R(q)| \leq 1$ in $\mathbb{C}^-$. Dit betekent dat er slechts een punt q_0 kan bestaan in $\mathbb{C}^-$ waarvoor $|R(q_0)| = 1$ als er een punt $q^* = i\,y^*$ bestaat waarvoor $|R(i\,y^*)| = 1$. Dit zou dan echter weer betekenen wegens het maximum modulus principe dat $R(q)$ constant is in $\mathbb{C}^-$, hetgeen we op voorhand uitsluiten. Derhalve geldt voor $q \in \mathbb{C}^-$ dat $|R(q)| < 1$, m.a.w. $R(q)$ is A-aanvaardbaar. ∎

Uit Lemma 4 volgt aldus duidelijk dat in Figuur 11.6 de enige ordester die correspondeert met een A-aanvaardbare stabiliteitsfunctie deze is in (b). Inderdaad, dit is de enige ordester waarvoor de doorsnede van (het witte gebied) $\mathcal{B}$ een lege doorsnede bezit met de imaginaire as.

De vier lemma's leiden nu tot de oplossing van de conjectuur van Ehle :

Stelling 11.4.4 *$\hat{R}_T^S(q)$ is slechts A-aanvaardbaar als $T - 2 \leq S \leq T$* □

Bewijs. Laat $\lfloor x \rfloor$ het grootste geheel getal voorstellen dat kleiner is of gelijk aan x en zij $R(q)$ een A-aanvaardbare benadering van de orde p. Dan zijn er wegens Lemma 1 minstens $\lfloor (p+1)/2 \rfloor$ vingers (waarbij een vinger van multipliciteit n telt als n vingers) die starten in $\mathbb{C}^-$. Wegens Lemma 4 kan geen enkele van deze vingers de imaginaire as snijden en bovendien garandeert Lemma 3 dat al deze vingers onbegrensd zijn. Dit betekent dat deze vingers samen de onbegrensde vinger vormen. Dit betekent ook dat deze onbegrensde vinger $\lfloor (p+1)/2 \rfloor - 1$ duale begrensde vingers uit $\mathbb{C}^-$ omvat, zodat wegens Lemma 3 minstens $\lfloor (p+1)/2 \rfloor - 1$ nulpunten in $\mathbb{C}^-$ liggen. Het aantal nulpunten van $R(q)$ bedraagt dus minstens $\lfloor (p+1)/2 \rfloor - 1$.
Stel nu dat $R(q) = \hat{R}_T^S(q)$, dan is $p = S + T$ en het aantal nulpunten bedraagt S. Uit $S \geq \lfloor (p+1)/2 \rfloor - 1$ volgt dan $2\,S + 2 \geq 2\,\lfloor (p+1)/2 \rfloor \geq p$, vermits

$$2\,\lfloor (p+1)/2 \rfloor = \begin{cases} p+1 & \text{als } p \text{ oneven is} \\ p & \text{als } p \text{ even is.} \end{cases}$$

Aldus is $2\,S + 2 \geq S + T$ of $S \geq T - 2$.
Gezien $S > T$ onmogelijk kan leiden tot A-aanvaardbaarheid, is de conjectuur van Ehle hiermee bewezen. ∎

11.5 Runge–Kutta-methoden voor stijve problemen

De vraag die nu nog rest is : bestaan er RKMn die bvb. A-stabiel of L-stabiel zijn ? Als we vertrekken van een RKM met Butcher-matrix

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

dan geeft dit, toegepast op de scalaire testvergelijking $y' = \lambda\, y$, $\lambda \in \mathbb{C}$, aanleiding tot de differentievergelijking

$$y_{n+1} = R(\hat{h})\, y_n$$

waarbij de stabiliteitsfunctie $R(\hat{h})$ een rationale functie is van $\hat{h} = \lambda\, h$. In paragraaf 11.1 werden twee vormen afgeleid voor $R(\hat{h})$:

$$R(\hat{h}) = 1 + h\, b^T\, (I_s - h\, A)^{-1}\, e\,, \tag{11.26}$$

$$R(\hat{h}) = \frac{\det\,[I_s - \hat{h}\, A + \hat{h}\, e\, b^T]}{\det\,[I_s - \hat{h}\, A]}\,, \tag{11.27}$$

waarbij $e = [1,\, 1,\, \ldots,\, 1]^T \in \mathbb{R}^s$. De methode zal A-stabiel of L-stabiel zijn zodra $R(\hat{h})$ A-aanvaardbaar of L-aanvaardbaar is. We weten reeds dat ERKMn niet in aanmerking komen. We zullen dus zoeken binnen de klasse van IRKMn of DIRKMn. Daartoe zouden we kunnen vertrekken van (11.26) of (11.27) en een beroep doen op resultaten van vorige paragraaf. Men kan echter ook enkele belangrijke resultaten afleiden zonder enig rekenwerk uit te voeren.

De Gaussische s-traps methode bezit orde $2\,s$, d.w.z. dat deze methode, toegepast op de testvergelijking $y' = \lambda\, y$, aanleiding geeft tot $y_{n+1} = R(\hat{h})\, y_n$ waarbij $R(\hat{h}) = \exp(\hat{h}) + \mathcal{O}(\hat{h}^{2s+1})$. Aldus is $R(\hat{h})$ een rationale benadering van $\exp(\hat{h})$ van orde $2\,s$. Uit (11.27) kan afgeleid worden dat zowel de teller als de noemer van $R(\hat{h})$ veeltermen zijn van graad hoogstens s. Anderzijds weten we dat er een unieke (s,s) rationale benadering bestaat van orde $2\,s$, nl. de Padé-benadering $\hat{R}^s_s(\hat{h})$. Hieruit volgt dus dat $R(\hat{h}) = \hat{R}^s_s(\hat{h})$, die wegens Stelling 11.4.1 A-aanvaardbaar is. *We kunnen aldus besluiten dat alle Gauss-methoden A-stabiel zijn en bovendien dat er A-stabiele IRKMn bestaan van willekeurige orde.*

Voor de andere klasses van IRKMn kunnen we gebruik maken van een aanpak volgens Dekker en Verwer. Deze aanpak buit de structuur van de matrix $A - e\, b^T$, die optreedt in (11.27), volledig uit.

Voor de Radau IA methode weten we dat de eerste kolom van de matrix A gevuld is met de waarde b_1. Dit betekent bijgevolg dat, als $+$ staat voor constante waarden en $\star$ voor een element dat hoogstens lineair is in $\hat{h}$:

$$A - e\, b^T = \begin{bmatrix} 0 & + & \ldots & + \\ 0 & + & \ldots & + \\ \vdots & & & \vdots \\ 0 & + & \ldots & + \\ 0 & + & \ldots & + \end{bmatrix}, \quad I_s - \hat{h}\,(A - e\, b^T) = \begin{bmatrix} 1 & \star & \ldots & \star \\ 0 & \star & \ldots & \star \\ \vdots & & & \vdots \\ 0 & \star & \ldots & \star \\ 0 & \star & \ldots & \star \end{bmatrix}.$$

De ontwikkeling van det $[I_s - \hat{h}\,(A - e\,b^T)]$ volgens de eerste kolom maakt duidelijk dat dit een polynoom van graad hoogstens $s-1$ kan zijn. Vermits anderzijds det $[I_s - \hat{h}\,A]$ hoogstens van graad s is in $\hat{h}$ en vermits de benadering van orde $2\,s-1$ is, kan $R(\hat{h})$ niets anders zijn dan de Padé-benadering $\hat{R}_s^{s-1}(\hat{h})$. M.b.v. Stelling 11.4.3 vinden we dat alle Radau IA methoden niet alleen A-stabiel maar zelfs L-stabiel zijn.

Voor de Radau IIA methode weten we dat de laatste rij van de matrix A gevuld is met de vector b. M.b.v. dezelfde notatie vinden we dan :

$$A - e\,b^T = \begin{bmatrix} + & + & \dots & + & + \\ + & + & \dots & + & + \\ \vdots & & & & \vdots \\ + & + & \dots & + & + \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad I_s - \hat{h}\,(A - e\,b^T) = \begin{bmatrix} \star & \star & \dots & \star & \star \\ \star & \star & \dots & \star & \star \\ \vdots & & & & \vdots \\ \star & \star & \dots & \star & \star \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Uit ontwikkeling van det $[I_s - \hat{h}\,(A - e\,b^T)]$ volgens de laatste rij volgt opnieuw dat dit een polynoom van graad hoogstens $s-1$ kan zijn. Vermits anderzijds det $[I_s - \hat{h}\,A]$ hoogstens van graad s is in $\hat{h}$ en vermits de benadering van orde $2\,s-1$ is, kan $R(\hat{h})$ opnieuw niets anders zijn dan de Padé-benadering $\hat{R}_s^{s-1}(\hat{h})$.

De Lobatto IIIA methode steunt op de veronderstelling dat de eerste rij van A alleen nullen bevat terwijl de laatste rij dezelfde is als de vector b. Dit betekent enerzijds dat det $[I_s - \hat{h}\,A]$ hoogstens van graad $s-1$ kan zijn en dat anderzijds $A - e\,b^T$ dezelfde structuur heeft als bij de Radau IIA methoden, en dus ook hoogstens van graad $s-1$ is. Vermits ze een benadering van orde $2\,s-2$ leveren, moet $R(\hat{h}) = \hat{R}_{s-1}^{s-1}(\hat{h})$, wat resulteert in A-stabiele methoden.

Voor Lobatto IIIB kunnen we vertrekken van het feit dat elk element in de eerste kolom van A gelijk is aan het element b_1 (waardoor $A - e\,b^T$ dezelfde structuur heeft als bij de Radau IA methoden en dus ook hoogstens van graad $s-1$ is), terwijl de laatste kolom van A alleen nullen bevat (waardoor det $[I_s - \hat{h}\,A]$ hoogstens van graad $s-1$ is). Aldus vinden we dezelfde conclusie als voor de Lobatto IIIA methoden.

Voor Lobatto IIIC methoden geldt dat de laatste rij van A de vector b is (zodat det $[I_s - \hat{h}\,A]$ hoogstens van graad s is), en dat de eerste kolom van A gevuld is met het element b_1. Dit leidt tot

$$A - e\,b^T = \begin{bmatrix} 0 & + & \dots & + & + \\ 0 & + & \dots & + & + \\ \vdots & & & & \vdots \\ 0 & + & \dots & + & + \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad I_s - \hat{h}\,(A - e\,b^T) = \begin{bmatrix} 1 & \star & \dots & \star & \star \\ 0 & \star & \dots & \star & \star \\ \vdots & & & & \vdots \\ 0 & \star & \dots & \star & \star \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

waaruit volgt dat det $[I_s - \hat{h}\,(A - e\,b^T)]$ hoogstens van graad $s-2$ kan zijn. Aldus wordt duidelijk dat $R(\hat{h}) = \hat{R}_s^{s-2}(\hat{h})$, wat, wegens Stelling 11.4.3, opnieuw resulteert in L-stabiele methoden.

De bovenstaande resultaten zijn samengevat in Tabel 11.2.

Tenslotte ook nog enkele resultaten i.v.m. SDIRKMn en SIRKMn. Daartoe vertrekken we van het volgende resultaat :

s-traps RKM	Orde	Stabiliteitsfunctie $R(\hat{h})$	Eigenschap inzake lineaire stabiliteit
Gauss	$2\,s$	$\hat{R}_s^s(\hat{h})$	A-stabiliteit
Radau IA, IIA	$2\,s-1$	$\hat{R}_s^{s-1}(\hat{h})$	L-stabiliteit
Lobatto IIIA, IIIB	$2\,s-2$	$\hat{R}_{s-1}^{s-1}(\hat{h})$	A-stabiliteit
Lobatto IIIC	$2\,s-2$	$\hat{R}_s^{s-2}(\hat{h})$	L-stabiliteit

Tabel 11.2:

Stelling 11.5.1 *Zij $R_2(q;\,\alpha,\,\beta)$ gedefinieerd door*

$$R_2(q;\,\alpha,\,\beta) = \frac{1+\frac{1}{2}\,(1-\alpha)\,q+\frac{1}{4}\,(\beta-\alpha)\,q^2}{1-\frac{1}{2}\,(1+\alpha)\,q+\frac{1}{4}\,(\beta+\alpha)\,q^2}\,,$$

dan is $R_2(q;\,\alpha,\,\beta)$ A-aanvaardbaar a.s.a. $\alpha \geq 0$ en $\beta \geq 0$ en anderzijds L-aanvaardbaar a.s.a. $\alpha = \beta > 0$. □

Voor algemene α en β heeft $R_2(q;\,\alpha,\,\beta)$ orde 2, orde 3 indien $\alpha \neq 0$ en $\beta = \frac{1}{3}$ en orde 4 (wat van $R_2(q;\,\alpha,\,\beta)$ een Padé-benadering maakt) voor $\alpha = 0$ en $\beta = \frac{1}{3}$.

We beschouwen vooreerst het SDIRK-paar

$$\begin{array}{c|cc} \frac{3\pm\sqrt{3}}{6} & \frac{3\pm\sqrt{3}}{6} & 0 \\ \frac{3\mp\sqrt{3}}{6} & \mp\frac{\sqrt{3}}{3} & \frac{3\pm\sqrt{3}}{6} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Het stabiliteitspolynoom voor beide methoden wordt gegeven door

$$R(\hat{h}) = \frac{1\mp\frac{\sqrt{3}}{3}\,\hat{h}-\frac{1\pm\sqrt{3}}{6}\,\hat{h}^2}{1-\frac{3\pm\sqrt{3}}{3}\,\hat{h}+\frac{2\pm\sqrt{3}}{6}\,\hat{h}^2} = R_2\left(\hat{h};\,1\pm 2\frac{\sqrt{3}}{3},\,\frac{1}{3}\right)\,,$$

wat volgens Stelling 11.5.1 betekent dat $R(\hat{h}) = R_2(\hat{h};\,\alpha,\,\beta)$ alleen A-aanvaardbaar is als $\alpha = 1+2\,\frac{\sqrt{3}}{3}$ en dat alleen de methode corresponderend met de bovenste van de alternerende tekens A-stabiel is.

Tenslotte beschouwen we ook de 2-traps SIRKM

$$\begin{array}{c|cc} (2-\sqrt{2})\,\mu & \frac{(4-\sqrt{2})\,\mu}{4} & \frac{(4-3\,\sqrt{2})\,\mu}{4} \\ (2+\sqrt{2})\,\mu & \frac{(4+3\,\sqrt{2})\,\mu}{4} & \frac{(4+\sqrt{2})\,\mu}{4} \\ \hline & \frac{4(1+\sqrt{2})\,\mu-\sqrt{2}}{8\,\mu} & \frac{4(1-\sqrt{2})\,\mu+\sqrt{2}}{8\,\mu} \end{array}$$

Deze methode bezit orde 3 indien $\mu = \frac{3\pm\sqrt{3}}{6}$ en anders orde 2. Gebruik makend van (11.9) vindt men

$$R(\hat{h}) = \frac{1 + (1 - 2\,\mu)\,\hat{h} + (\mu^2 - 2\,\mu + \frac{1}{2})\,\hat{h}^2}{1 - 2\,\mu\,\hat{h} + \mu^2\,\hat{h}^2} = R_2(\hat{h};\, 4\,\mu - 1,\, (2\,\mu - 1)^2)\,.$$

Volgens Stelling 11.5.1 betekent dit dat $R(\hat{h}) = R_2(\hat{h};\, \alpha,\, \beta)$ A-aanvaardbaar is voor $\mu \geq \frac{1}{4}$. De methode van orde 3 waarvoor $\mu = \frac{3+\sqrt{3}}{6}$ is aldus A-stabiel. Uit Stelling 11.5.1 volgt tenslotte ook nog dat deze familie van methoden een paar tweede orde L-stabiele methoden bevat voor $\mu = \frac{2\pm\sqrt{2}}{2}$.

Het moge aldus duidelijk zijn dat we geen grote problemen ondervinden om IRKMn of DIRKMn te vinden die A- of L-stabiel zijn. Elk van deze methoden kan, zonder al te veel problemen, gebruikt worden in een code.