Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

COLLOQUIUM STIJVE DIFFERENTIAALVERGELIJKINGEN

DEEL 1

DOOR T.J.DEKKER, P.W.HEMKER EN P.J. VAN DER HOUWEN

MC SYLLABUS 15.1

MATHEMATISCH CENTRUM AMSTERDAM 1972

Inhoud

Voorwoord

Het colloquium 'stijve differentiaalvergelijkingen' had tot doel een overzicht te geven van het onderzoek, verricht door de werkgroep 'stijve differentiaalvergelijkingen' van het Mathematisch Centrum. De meeste aspecten van dit onderzoek zijn in het colloquium naar voren gebracht en worden weergegeven in deze syllabus.

Het onderwerp waaraan het colloquium gewijd was, betrof het numeriek oplossen van beginwaardeproblemen voor gewone differentiaalvergelijkingen. Hierbij werden vooral methoden behandeld welke geschikt zijn voor het oplossen van stijve differentiaalvergelijkingen.

In het eerste deel van deze syllabus, waarin de eerste vier bijdragen aan het colloquium zijn opgenomen, wordt een inleiding tot de problematiek, een historisch overzicht en de behandeling van eenstaps- en meerstapsmethoden gegeven.

In het tweede gedeelte zullen meer speciale onderwerpen aan de orde komen zoals exponentieel aangepaste methoden, stapkeuze strategieën en het schatten van parameters in differentiaalvergelijkingen.

P.W.H.

1. Inleiding

Het gebruik om verschijnselen te beschrijven met behulp van differentiaalvergelijkingen is de laatste tientallen jaren in verschillende takken van biologisch en biochemisch onderzoek ingeburgerd. Als enkele van de belangrijkste gebieden van onderzoek waar dit gebeurt kunnen we noemen de populatie-dynamica, de tracer-kinetica, de enzymkinetica, de eiwitsynthese en de morphogenese. Deze opsomming is verre van volledig en telkens worden nieuwe, waaronder zeer ingenieuze, fysiologische en biologische modellen beschreven in termen van differentiaalvergelijkingen (zie bijv. het tijdschrift Mathematical Biosciences).
We zullen hier op deze onderwerpen zelf niet verder ingaan; we zullen ons echter, aan de hand van enkele eenvoudige praktijkvoorbeelden, voornamelijk concentreren op het numeriek oplossen van de differentiaalvergelijkingen en wel in het bijzonder op die gevallen waar klassieke oplossingsmethoden geen bevredigende resultaten leveren.

Aangezien de differentiaalvergelijkingen die in de biologische disciplines (de tracer-kinetica uitgezonderd) verschijnen, bijna allen niet-lineair en van een zodanige vorm zijn dat een analytische oplossing niet beschikbaar is, doet de noodzaak van numerieke methoden zich direkt gevoelen. We willen er hier al de nadruk op leggen dat het bestaan van goede numerieke methoden niet een analytisch onderzoek overbodig maakt. Enig analytisch onderzoek zal altijd nodig zijn om de juiste numerieke methode te kunnen vinden en om de betrekkelijkheid van het numeriek verkregen resultaat in te zien.

We zullen ons hier beperken tot beginwaardeproblemen voor gewone differentiaalvergelijkingen - de belangrijkste klasse van differentiaalvergelijkingen in biomathematisch onderzoek - en wel in het bijzonder tot stijve differentiaalvergelijkingen.

Stijve differentiaalvergelijkingen

Vele stelsels gewone differentiaalvergelijkingen, welke in de praktijk opgesteld worden, hebben de eigenschap dat de oplossing zowel snel varierende als langzaam varierende componenten bevat. Men kan zich een

stelsel differentiaalvergelijkingen voorstellen dat een electronische schakeling beschrijft waarbij het inschakelverschijnsel aanzienlijk sneller verloopt dan de uiteindelijke werking van de schakeling. Of men stelt zich een stelsel differentiaalvergelijkingen voor dat een aantal gekoppelde chemische reacties beschrijft waarbij sommige reacties vele malen sneller verlopen dan andere. We noemen, naar Curtiss en Hirschfelder [1952], een stelsel differentiaalvergelijkingen met deze eigenschap stijf ("stiff") omdat ook een vergelijking, die de werking beschrijft van een eenvoudig mechanisch model met een stijve veer, tot deze categorie behoort.

We kunnen voor biologen het begrip "stijve vergelijking" ook verklaren aan de hand van het begrip "epigenetisch landschap" van Waddington [1957]. Laat de toestand van het biologische systeem dat we beschrijven vastgelegd worden door twee toestandsvariabelen X en Y. We denken ons het "epigenetisch landschap" als een reële functie van X en Y. Een biologisch systeem dat zich in een bepaalde toestand (X_0,Y_0) bevindt, zal zich bewegen naar die toestanden (X,Y) die bereikt worden door langs het pad met de snelste afdaling te lopen. De differentiaalvergelijkingen vertonen een stijf karakter op die plaatsen waar de dalen in het epigenetisch landschap smal zijn.

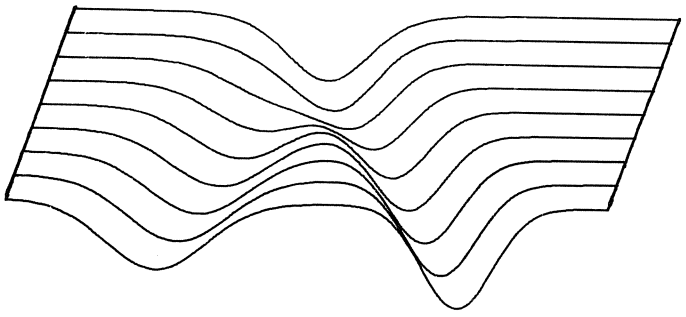


fig. 1.1 Epigenetisch landschap.

We kunnen hetzelfde illustreren met behulp van een richtingsveld

$$\frac{dy}{dx} = f(x,y).$$

Als voorbeeld nemen we in figuur 1.2 de differentiaalvergelijking

$$\frac{dy}{dx} = -2.5\ y + \frac{(5x+3)}{(x+1)^2} \quad .$$

Het oplossen van de differentiaalvergelijking komt overeen met het vinden van een baankromme in het richtingsveld. Een differentiaalvergelijking vertoont een stijf karakter wanneer alle oplossingen snel convergeren naar een bepaalde verzameling langzaam varierende oplossingen (de asymptotische oplossingen). Een aantal van deze baankrommen van het richtingsveld in figuur 1.2 is getekend in figuur 1.3 waarop ook de asymptotische oplossing duidelijk te zien is.

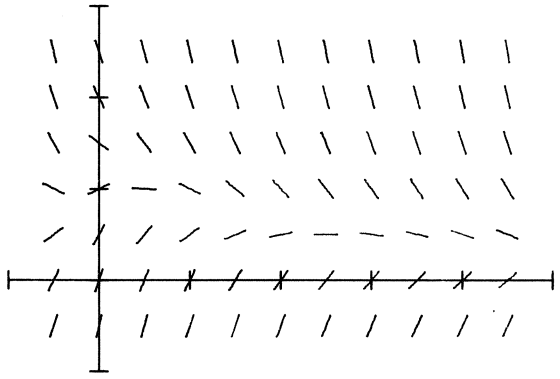


fig. 1.2 Het richtingsveld voor de differentiaalvergelijking $\frac{dy}{dx} = -2.5\ y + \frac{5x+3}{(x+1)^2}$.

Dergelijke vergelijkingen, die verschijnselen beschrijven met ver uiteenliggende tijdconstanten, leveren bij numerieke integratie volgens standaardmethoden moeilijkheden op. Dit vindt zijn oorzaak in het feit dat voor het verkrijgen van een numeriek stabiel proces tijdstappen worden vereist met een zodanige orde van grootte dat de snelst varierende compo-

nent gevolgd kan worden, zodat het volgen van de langzame component een bijzonder grote hoeveelheid rekenwerk vergt. We mogen opmerken dat zeer veel verschillende begrippen de naam 'stabiliteit' dragen; het hier ter sprake gekomen begrip numerieke stabiliteit zullen we hierna toelichten.

Omdat de stijve vergelijkingen - mits stabiel - een hyperstabiel gedrag vertonen - dit houdt in dat het verschijnsel zich na verloop van tijd in belangrijke mate onafhankelijk van het inschakelverschijnsel zal gedragen - ligt het voor de hand te veronderstellen dat er algorithmen te vinden zijn die de genoemde moeilijkheid grotendeels ondervangen. Daar het bovendien duidelijk zal zijn dat er geen scherpe grens getrokken kan worden tussen stijve en niet-stijve differentiaalvergelijkingen, blijft de moeilijkheid bestaan methoden te construeren die de gunstige eigenschappen van standaardmethoden combineren met eigenschappen van methoden die geschikt zijn voor stijve differentiaalvergelijkingen.

Kwantitatieve beschrijving van stijfheid

We zullen nu een manier aangeven om de stijfheid van een stelsel differentiaalvergelijkingen kwantitatief te beschrijven. Zij gegeven een stelsel differentiaalvergelijkingen in vectorvorm:

$$\frac{d}{dx}\ \vec{y}(x) = \vec{f}(x,\vec{y}). \tag{1.1}$$

Wanneer de vectorfunctie $\vec{f}$ differentieerbaar is naar $\vec{y}$ kunnen we $\vec{f}$ locaal lineariseren ter plaatse $\vec{y}_0$; we schrijven

$$\frac{d}{dx}\ \vec{y}(x) = \vec{H}(x) + \vec{\vec{J}}(x,\vec{y}_0)(\vec{y}-\vec{y}_0) + \ldots \tag{1.2}$$

waarin $\vec{H}(x) = \vec{f}(x,\vec{y}_0)$ een vector is en $\vec{\vec{J}}(x,\vec{y}_0) = (\frac{\partial f_i}{\partial y_j})_{\vec{y}=\vec{y}_0}$ de Jacobiaan ter plaatse $\vec{y}_0$ voorstelt.
Hoewel het gedrag van de oplossing natuurlijk ook afhangt van de term $\vec{H}(x)$, wordt - mits $\vec{H}(x)$ langzaam varieert met x - een goede kwantitatieve beschrijving van de stijfheid verkregen door de eigenwaarden van de Jacobiaan $\vec{\vec{J}}$ in het complexe vlak te localiseren. We kunnen ons dit als volgt

voorstellen. Ter plaatse $(x_0,\vec{y}_0)$ is (1.1) te benaderen door

$$\frac{d}{dx}\vec{y}(x) = \vec{H}(x_0) + \vec{\vec{J}}(x_0,\vec{y}_0)(\vec{y}-\vec{y}_0) \quad . \tag{1.3}$$

Wanneer we aannemen dat de eigenwaarden van $\vec{\vec{J}}$ allen verschillend zijn zal de locaal analytische oplossing zich derhalve laten schrijven als

$$\vec{y}(x) - \vec{y}(x_0) = \vec{b} + \sum_i c_i \vec{E}_i \, e^{\lambda_i(x-x_0)} ,$$

waarin $\{\lambda_i\}$ en $\{\vec{E}_i\}$ resp. de eigenwaarden en eigenvectoren van $\vec{\vec{J}}$ zijn en waarin $\vec{b}$ en $\{c_i\}$ bepaald worden door de lineaire stelsels

$$\vec{\vec{J}}\,\vec{b} + \vec{H}(x_0) = 0 \text{ resp. } \quad 0 = \vec{b} + \sum_i c_i \vec{E}_i \quad .$$

Aan deze locale beschouwing zien we dat het tijdsafhankelijke gedrag van de oplossing in eerste instantie bepaald wordt door de eigenwaarden (de tijdconstanten) en door het gedrag van $\vec{H}(x)$.

Bij een stabiel stelsel differentiaalvergelijkingen bevinden alle eigenwaarden zich in het halfvlak Re $\lambda_i \leq 0$, terwijl voor een stijf stelsel de eigenwaarden zowel in de omgeving van de oorsprong als verspreid over het halfvlak Re $\lambda_i \leq 0$ liggen.

Numerieke stabiliteit

Zoals we in het voorafgaande opmerkten is het knelpunt bij het oplossen van stijve vergelijkingen de numerieke stabiliteit. Een rekenproces heet numeriek instabiel wanneer een door het proces geïntroduceerde fout (bijvoorbeeld een afrondingsfout) tijdens de berekening systematisch toeneemt en daardoor het resultaat van de berekening overvleugelt. Een proces heet numeriek stabiel als een eenmaal geïntroduceerde fout afneemt.

We zullen het begrip numerieke stabiliteit voor een methode om differentiaalvergelijkingen op te lossen toelichten aan de hand van een zeer eenvoudig - maar representatief - voorbeeld.

We lossen de differentiaalvergelijking

$$dy/dx = f(x,y) = \lambda y + g(x) \quad (\lambda < 0)$$

op met behulp van de methode van Euler.
Na keuze van een staplengte $h > 0$ willen we, uitgaande van het punt $y(x)$, de waarde van $y(x+h)$ berekenen.
Volgens de methode van Euler:

$$\begin{aligned} y(x+h) &= y(x) + h * f(x,y(x)) \\ &= y(x) + h\lambda y(x) + h\, g(x) \\ &= (1+h\lambda)\, y(x) + h\, g(x). \end{aligned}$$

Wanneer de reeds berekende waarde $y(x)$ bestaat uit een juiste waarde $\tilde{y}(x)$ en fout ε:

$$y(x) = \tilde{y}(x) + \varepsilon$$

dan geeft deze fout ε aanleiding tot een fout in de berekende waarde $y(x+h)$ ter grootte van $(1+h\lambda)\varepsilon$. Immers

$$\begin{aligned} y(x+h) &= (1+h\lambda)(\tilde{y}(x)+\varepsilon) + h\, g(x) \\ &= (1+h\lambda)\, \tilde{y}(x) + h\, g(x) + (1+h\lambda)\varepsilon \\ &= \tilde{y}(x+h) + (1+h\lambda)\varepsilon\ . \end{aligned}$$

De eis dat een eenmaal geïntroduceerde fout kleiner moet worden, komt overeen met de eis

$$|(1+\lambda h)\,\varepsilon| < |\varepsilon| \quad \text{ofwel} \quad h < \left|\frac{2}{\lambda}\right| \quad . \tag{1.4}$$

We zien dat de eis van numerieke stabiliteit ons een bovengrens voor de staplengte geeft. In figuur 1.3 hebben we enige integratie-stappen getekend voor het geval $\lambda = -2.5$ en $h = 1$.

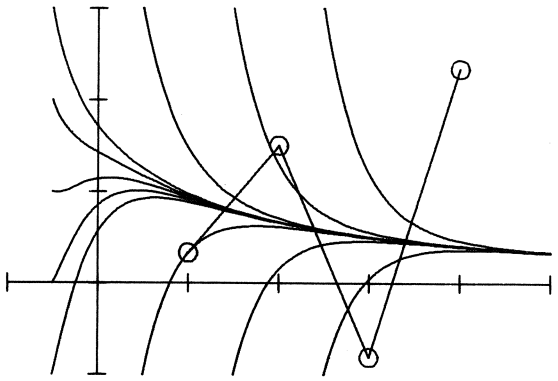


fig. 1.3 De Eulermethode, $\lambda = -2.5$ en $h = 1$.

We kunnen ook laten zien dat er eenvoudige methoden bestaan waarbij numerieke stabiliteit geen grens aan de staplengte stelt. Deze methoden hebben daarentegen het nadeel dat bij elke stap een (in het algemeen niet-lineaire) vergelijking of stelsel vergelijkingen moet worden opgelost. Als voorbeeld lossen we dezelfde differentiaalvergelijking

$$dy/dx = f(x,y) = \lambda y + g(x) \qquad (\lambda < 0)$$

nu op met de backward-Euler methode:

$$\begin{aligned} y(x+h) &= y(x) + h.f(x+h, y(x+h)) \\ &= y(x) + h\lambda\, y(x+h) + h.g(x+h) \end{aligned}$$

$$y(x+h) = \frac{y(x) + h.g(x+h)}{1 - \lambda h} \quad .$$

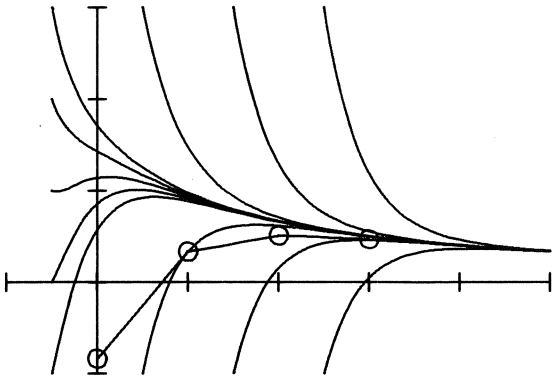


fig. 1.4 De backward-Euler methode, $\lambda = -2.5$ en $h = 1$.

Een fout ε in de grootheid $y(x)$ veroorzaakt nu een fout $\varepsilon/(1-\lambda h)$ in $y(x+h)$. De eis voor numerieke stabiliteit luidt nu

$$|1-\lambda h| > 1 \ . \tag{1.5}$$

Voor een stabiel probleem ($\lambda < 0$) wordt hier geen grens aan de staplengte gesteld.

De vorm van de stabiliteitsvoorwaarden (1.4) en (1.5) geeft ook enige rechtvaardiging voor het feit dat we bij de kwantitatieve beschrijving van stijfheid de inhomogene term $H(x)$ buiten beschouwing gelaten hebben.

Een laatste opmerking die hier wellicht over de foutenopbouw gemaakt moet worden is de volgende. De beschreven methoden om beginwaardeproblemen op te lossen geven een voorschrift om stap voor stap $y(x)$ te bepalen. Laat ε^*_{n-1} de totale fout zijn welke in de berekening van $y(t_{n-1})$ is opgetreden en laat de bijdrage van ε^*_{n-1} tot ε^*_n gegeven worden door $\alpha_n \varepsilon^*_{n-1}$ (α_n is de amplificatie factor). Aangezien bij elke stap bovendien telkens een nieuwe fout ε_n geïntroduceerd wordt zal gelden

$$\varepsilon^*_n = \alpha_n \varepsilon^*_{n-1} + \varepsilon_n \ .$$

Voor de amplificatiefactor hebben we geeist $|\alpha_n| \leq A < 1$. Wanneer we aannemen dat $|\varepsilon_n|$ begrensd is met een bovengrens E kunnen we nu aantonen dat de totale fout in de berekening begrensd blijft met de bovengrens $\frac{E}{1-A}$. Voor iedere n geldt namelijk

$$
\begin{aligned}
|\varepsilon_n^*| &\leq |\varepsilon_n| + |\alpha_n||\varepsilon_{n-1}^*| \leq |\varepsilon_n| + |\alpha_n| \,(|\varepsilon_{n-1}|+|\alpha_{n-1}||\varepsilon_{n-2}^*|) \\
&\leq |\varepsilon_n| + |\alpha_n||\varepsilon_{n-1}| + |\alpha_n||\alpha_{n-1}||\varepsilon_{n-2}| + \ldots \\
&\leq E + A \,.\, E + A^2 \,.\, E + \ldots \\
&= E\,(1+A+A^2+\ldots) = \frac{E}{1-A} \,.
\end{aligned}
$$

Analoge berekeningen laten zien dat A = 1 tot een lineaire foutenopbouw en dat A > 1 tot een exponentiële foutenopbouw aanleiding kan geven.

Een voorbeeld uit de enzymkinetica

We kunnen nu een probleem uit de praktijk kiezen en een voorbeeld geven van een wiskundige analyse. Uit de ruime keus van wiskundige problemen die de biomathematica ons biedt, hebben we een enkel biochemisch probleem gelicht (1) omdat het een systeem beschrijft dat telkens, bij het simuleren van biochemische systemen, als deelsysteem voorkomt en (2) omdat het enige eigenschappen heeft die een nadere analyse waard zijn zoals niet-lineariteit en stijf gedrag.

We behandelen een eenvoudige enzymreactie van het Michaelis-Menten type. Deze chemische reactie is van de vorm

$$
E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C \xrightarrow{k_2} E + P
$$

Dit schema beschrijft de reacties van een enzymmolecuul E dat zich reversibel met een substraatmolecuul S tot een enzym-substraat complex C bindt, terwijl het gevormde complex irreversibel omgezet kan worden in het oorspronkelijke enzym E een een produkt P. De reactieconstanten van de ver-

schillende deelreacties worden aangegeven met k_1, k_{-1} en k_2. Als regel geldt in dit soort reacties dat de concentratie van het enzym klein is ten opzichte van de concentratie van het substraat terwijl bovendien in vele gevallen geldt $k_{-1} >> k_2$. De wet van de massawerking stelt ons in staat om het gedrag van de reactie in de loop van de tijd te beschrijven *:

$$(1.6) \qquad \left.\begin{aligned} \frac{d}{dt} S &= -k_1(E_0-C)S + k_{-1} C \\ \frac{d}{dt} C &= k_1(E_0-C)S - (k_2+k_{-1})C \end{aligned}\right\}$$

Als beginvoorwaarden nemen we $S_{t=0} = S_0$, $E_{t=0} = E_0$ en $C_{t=0} = 0$. Om de notatie zo eenvoudig mogelijk te houden, zullen we met de volgende substituties het probleem in een dimensieloze vorm schrijven.

$$s(t) = S/S_0 \qquad c(t) = C/E_0$$
$$\varepsilon = E_0/S_0 \qquad \tau = t\, k_1\, E_0$$
$$p = (k_2+k_{-1})/(k_1 S_0) \qquad q = k_{-1}/(k_1 S_0)$$

p is de dimensieloze Michaelis-constante.

Als resultaat krijgen we

$$(1.7) \qquad \left.\begin{aligned} \frac{ds}{d\tau} &= -(1-c)s + qc \\ \varepsilon \frac{dc}{d\tau} &= (1-c)s - pc \end{aligned}\right\}$$
$$s(0) = 1; \quad c(0) = 0 \quad .$$

We merken op dat van de verschillende grootheden nu het volgende bekend is.

$$\varepsilon, \tau, q > 0 \ , \quad p > q$$
$$0 \leq s, c \leq 1$$

Als regel geldt $\varepsilon << 1$ (een kleine parameter) en dikwijls $0 < p-q << q$.

---

*) N.B. De letters E, S, C en P worden hier zowel gebruikt om de chemische verbindingen als om de concentraties van de betreffende verbindingen aan te geven.

De numerieke waarden van p, q en ε kunnen overigens van geval tot geval sterk uiteenlopen [o.a. Briggs en Haldane (1925)].

Om aan te tonen dat het stelsel (1.7) een typisch stijf karakter heeft berekenen we de Jacobiaan van het stelsel.

$$J = \begin{pmatrix} -(1-c) & q+s \\ (1-c)/\varepsilon & -(p+s)/\varepsilon \end{pmatrix}.$$

Voor het onderzoek naar de eigenwaarden λ_M en λ_m van deze Jacobiaan berekenen we het spoor en de determinant:

$$sp(J) = -[(1-c) + \frac{p+s}{\varepsilon}]$$

$$det(J) = \frac{(p-q)(1-c)}{\varepsilon} \quad .$$

Hieruit volgt direkt $\lambda_M < \lambda_m < 0$ en

$$2(1 + \frac{\lambda_M}{\lambda_m}) \leq \frac{(\lambda_m+\lambda_M)^2}{\lambda_m.\lambda_M} = \frac{(sp(J))^2}{det(J)} > (\frac{p+s}{\varepsilon})^2/(\frac{p-q}{\varepsilon}) = \frac{(p+s)^2}{\varepsilon(p-q)} .$$

Hieruit blijkt dat beide eigenwaarden negatief zijn terwijl de verhouding van de absolute waarden groot is, hetgeen karakteristiek is voor stijve vergelijkingen.

Daar de biochemische onderzoeker bij het simuleren van deze soort enzymreacties in het algemeen slechts Euler en standaard Runge-Kutta methoden tot zijn beschikking had, is het op grond van het bovenstaande niet verwonderlijk dat deze stijve differentiaalvergelijkingen hem dikwijls voor grote moeilijkheden hebben gesteld. Deze moeilijkheden die zich manifesteerden in extreem lange rekentijden, werden in voorkomende gevallen niet opgelost door gebruik te maken van betere integratiemethoden maar omzeild door het model van de beschouwde reacties te wijzigen [D. Garfinkel and B. Hess (1964), J.G. Reich (1968). D. Garfinkel e.a. (1970)]

Een verband met singuliere storingsrekening

In de enzymkinetica [o.a. M. Dixon and E. Webb] zijn enkele benaderende oplossingen bekend voor het stelsel (1.6) namelijk de Briggs-Haldane formule [Briggs en Haldane (1925)] en de formule van Gutfreund [Gutfreund (1965)]. Een gebruikelijke wijze om benaderingen te vinden voor stelsels differentiaalvergelijkingen in gevallen waar een hoogste afgeleide wordt vermenigvuldigd met een kleine parameter, zoals in (1.7), wordt gevonden in de theorie van de singuliere storingsproblemen [o.a. J.D. Cole (1968), Heineken et al. (1967)].
We zullen in het volgende laten zien hoe deze theorie de reeds bij biochemici bekende formules als een bijzonder geval van een eerste benadering doet uitkomen.

We beschouwen het stelsel (1.7) en proberen een oplossing te vinden welke asymptotisch juist is voor $\varepsilon \to 0$. Hiertoe stellen we in eerste instantie $\varepsilon = 0$ zodat we krijgen

$$\left.\begin{aligned} \frac{ds}{d\tau} &= -(1-c)s + qc \\ 0 &= (1-c)s - pc \end{aligned}\right\}$$

$$s(0) = 1 \quad ; \quad c(0) = 0$$

Wanneer we dit stelsel oplossen krijgen we

$$c = \frac{s}{s+p}$$

(d.i. de dimensieloze Briggs-Haldane formule) en

$$\frac{ds}{d\tau} = -(p-q)\,\frac{s}{s+p} \quad .$$

Deze laatste differentiaalvergelijking laat een impliciete oplossing voor $s(\tau)$ toe:

$$s(\tau) + p\,\ln(s(\tau)) + (p-q)\tau = 1. \tag{1.8}$$

Het is duidelijk dat op deze wijze niet voldaan kan worden aan de randvoorwaarde $c(0) = 0$; daartoe voeren we bij $\tau = 0$ een locale coördinaat in $\zeta = \tau/\varepsilon$. Wanneer we (1.7) in deze variabele uitdrukken krijgen we de beschrijving van het inschakelverschijnsel:

$$(1.9)\qquad \left.\begin{aligned} \frac{ds}{d\zeta} &= -\varepsilon(1-c)s + \varepsilon q c \\ \frac{dc}{d\zeta} &= (1-c) - pc \end{aligned}\right\}$$
$$s(0) = 1;\ c(0) = 0 \quad .$$

Nemen we weer $\varepsilon = 0$ dan krijgen we

$$\left.\begin{aligned} \frac{ds}{d\zeta} &= 0 \\ \frac{dc}{d\zeta} &= (1-c)s - pc \end{aligned}\right\}$$

waarvan de oplossing luidt:

$$s(\zeta) = 1$$
$$c(\zeta) = \frac{1}{1+p}\left[1 - e^{-(1+p)\zeta}\right]$$

(d.i. de dimensieloze vorm van de formule van Gutfreund).
Nu geldt

$$\lim_{\zeta\to\infty} s(\zeta) = 1 = \lim_{\tau\to 0} s(\tau)$$

en

$$\lim_{\zeta\to\infty} c(\zeta) = \frac{1}{1+p} = \lim_{\tau\to 0} c(\tau)$$

zodat aan de "matching conditions" [zie bijv. J.D. Cole (1968)] is voldaan en we een asymptotische oplossing $O(\varepsilon)$ van (1.7) op een willekeurig traject $[0,T]$ kunnen geven.

$$(1.10)\qquad \begin{cases} s(\tau) = s(\tau) \text{ gedefinieerd door (1.8)} \\ c(\tau) = \dfrac{s(\tau)}{s(\tau)+p} - \dfrac{1}{1+p}\, e^{-(1+p)\tau/\varepsilon} \end{cases}$$

De uniforme geldigheid van deze oplossing op $[0,\infty)$ kan eenvoudig worden aangetoond. De theorie van de singuliere storingsproblemen stelt ons ook in staat hogere orde benaderingen te vinden. We zullen echter hierop niet verder ingaan.

Voor diegenen die de impliciete definitie voor $s(\tau)$ (1.8) een onbevredigend resultaat vinden, mogen we opmerken dat juist met behulp van een computer op eenvoudige wijze voor elke waarde van τ de bijbehorende waarde voor $s(\tau)$ berekend kan worden.

Uit het voorafgaande is duidelijk geworden dat voor degeen die een concreet probleem numeriek wil oplossen dikwijls een groot aantal geheel verschillende methoden ter beschikking staat. Nu eens zal de ene methode beter zijn, dan weer een andere. Zoekt men voor een gecompliceerd probleem een optimale oplossingsmethode dan zal men niet moeten terugschrikken voor een grondige analyse.

Wanneer men soms in de literatuur [bijv. Garfinkel (1968)] aanbevelingen leest voor computerprogramma's waarbij van de gebruiker geen enkel wiskundig of numeriek inzicht geeist wordt, is enige scepsis gerechtvaardigd en zal men toch in de eerste plaats verbaasd moeten zijn over het feit dat dergelijke programma's soms blijken te voldoen.

Wij stellen ons voor in dit colloquium een beschrijving te geven van een aantal methoden voor het oplossen van beginwaardeproblemen zodat een gebruiker eventuele moeilijkheden zal kunnen onderkennen en een methode geschikt voor zijn concrete probleem zal kunnen uitkiezen.

Literatuur

Briggs, G.E. and Haldane, J.B.S.
A note on the kinetics of enzyme action.
Biochem. J. 19(1925) 338.

Cole, J.D.
Perturbation methods in applied mathematics.
Blaisdell Publ. Comp. (1968).

Curtiss, C.F. and Hirschfelder, J.O.
Integration of stiff equations.
Proc. Nat. Acad. Sc. U.S. 38(1952) 235.

Dixon, M. and Webb, E.
Enzymes.
Longmans Green and Co. (1965).

Garfinkel, D. and Hess, B.
Metabolic control mechanisms.
A detailed computer model of the glycolytic pathway in ascites cells.
J. Biol. Chem. 239(1964) 971.

Garfinkel, D.
Construction of biochemical computer models.
in: Computing techniques in Biochemistry.
(J.H. Ottaway ed.) FEBS Letters Vol. 2(1968) supp. p.9

Garfinkel, D., Garfinkel, L., Pring, M., Green, S.B. and Change, B.
Computer applications to biochemical kinetics.
Ann. Rev. Biochem. 39(1970) 473.

Gutfreund, H.
An introduction to the study of enzymes.
Blackwell Scientific Publications (1965).

Heineken, F.G., Tsuchiya and Aris, R.
On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics.
Math. Biosc. 1(1967) 95.

Reich, J.G.
Aims, applications and achievements.
in: Computing techniques in Biochemistry
(J.H. Ottaway ed.) FEBS Letters Vol.2(1968) Supp. p.3.

Rosen, R.
Dynamical system theory in biology.
Wiley-Interscience (1970).

Waddington, C.H.
The strategy of the genes.
Ruskinhouse & Unwin Ltd. (1957).

2. Historisch overzicht

We schrijven een beginwaarde-probleem bestaande uit s eerste-orde vergelijkingen en een bijbehorende beginvoorwaarde in de volgende algemene vorm

$$\frac{d}{dx} y(x) = f(x,y(x)), \tag{2.1}$$

$$y(x_0) = y_0. \tag{2.2}$$

Hierin is $x \in R$ de onafhankelijke variabele, het rechterlid $f \in R \times R^s \to R^s$ een gegeven functie (met componenten $f^1, \ldots, f^s$), $y_0 \in R^s$ een gegeven vector van beginwaarden $y_0^1, \ldots, y_0^s$ voor het startpunt $x_0 \in R$, en $y \in R \to R^s$ de gezochte onbekende vector van functies $y^1, \ldots, y^s$.
Stelsels vergelijkingen van hogere orde kan men gemakkelijk tot stelsels van de eerste orde herleiden en worden daarom hier niet apart behandeld. De Jacobiaan J van het stelsel is de matrix der partiële afgeleiden

$$J_{ij} = f^i_{y^j}(x,y(x)), \qquad i,j = 1, \ldots, s. \tag{2.3}$$

Voor het verkrijgen van overzichtelijke integratieformules is het vaak handig het stelsel *autonoom* te maken, d.w.z. de onafhankelijke variabelen (als 0^e component) in de vector van onbekende functies op te nemen. Stellen we dienovereenkomstig

$$f_0 \equiv 1, \quad y_0^0 = x_0, \tag{2.4}$$

dan krijgt het beginwaarde-probleem de gedaante

$$\frac{d}{dx} y(x) = f(y(x)), \tag{2.5}$$

$$y(x) = y_0, \tag{2.6}$$

waarbij $f \in R^{s+1} \to R^{s+1}$, $y_0 \in R^{s+1}$ en $y \in R \to R^{s+1}$, terwijl uit (2.3) volgt

$$y^0(x) \equiv x.$$

De Jacobiaan J van dit stelsel heeft, behalve de elementen gegeven door (2.3), blijkbaar een extra 0^e rij en 0^e kolom, waarvoor geldt:

$$
\begin{aligned}
&J_{0j} = 0, && j = 0, \ldots, s, \\
&J_{i0} = f_x^i(x,y(x)), && i = 1, \ldots, s.
\end{aligned}
\tag{2.7}
$$

Wij beschouwen numerieke oplossingsmethoden gebaseerd op discretisatie van de x-as. Slechts voor een discrete verzameling van argument-waarden $\{x_0,x_1,x_2,\ldots\}$ worden benaderde waarden van de oplossing y berekend. Dit geschiedt stapsgewijs als volgt. In de n^e stap ($n \geq 0$) zijn benaderde waarden y_i van $y(x_i)$ gegeven voor $i = 0, \ldots, n$ en wordt een benaderding y_{n+1} van $y(x_{n+1})$ berekend voor zekere geschikt gekozen x_{n+1}.
De waarde $h = h_n \stackrel{=}{\text{def}} x_{n+1} - x_n$ heet de n^e *staplengte*.

Formules voor het berekenen van y_{n+1} worden onderscheiden in *eenstaps-formules* en *meerstaps-formules*, al naar gelang ze van de gegeven argumenten en bijbehorende functiewaarden alleen de laatste (x_n,y_n) of ook vroegere waarden $(x_i,y_i$ voor $i < n)$ gebruiken.

Bovendien worden de formules onderscheiden in *expliciete* en *impliciete* formules, welke laatste voor het berekenen van y_{n+1} de oplossing van een algebraïsch of transcendent stelsel vergelijkingen vergen.

Nauwkeurigheid en efficiëntie

Vóór de komst van rekenautomaten streefde men vooral naar het bereiken van een behoorlijke precisie met behulp van eenvoudig rekenwerk. Om deze reden waren in die tijd de meerstapsformules (met constante staplengte) favoriet. De stabiliteitseis was weliswaar minder stringent dan voor de tegenwoordig niet ongebruiklijke lange berekeningen, maar kon voor de klasse der meerstapsformules niet geheel verwaarloosd worden. Formules die niet stabiel zijn voor kleine positieve staplengten terwijl het stelsel differentiaalvergelijkingen een stabiele Jacobiaan bezit, d.w.z. dat alle eigenwaarden ervan een niet-positief reëel deel hebben, waren niet of slechts met omzichtigheid te gebruiken.

Om een redelijke stabiliteit te waarborgen gebruikte men meestal een

combinatie van een expliciete en een impliciete formule (predictor-correctormethode), waarbij de impliciete formule zo nodig iteratief werd toegepast.

Voor het oplossen van differentiaalvergelijkingen met behulp van rekenautomaten werd en wordt bovendien vaak gebruik gemaakt van (expliciete) Runge-Kuttaformules, die weliswaar meer rekenwerk vergen, maar het grote voordeel hebben dat de staplengte zonder moeite gevarieerd kan worden, zodat men deze aan het gedrag van de oplossing kan aanpassen.

Een andere methode, waarin eveneens de staplengte gemakkelijk gevarieerd kan worden, is die van Nordsieck (1962). Deze methode is gebaseerd op een meerpunts-polynoombenadering van de oplossing, waarvan in elke stap de afgeleiden worden bijgehouden.

De Runge-Kutta-formules zijn gebaseerd op een Taylorreeksontwikkeling, waarbij de afgeleiden van de rechterlid-functie f worden benaderd door bepaalde lineaire combinaties van waarden van f voor geschikt gekozen argumentwaarden.

Gill (1951) ontwierp een 4^e orde Runge-Kutta-formule, die minder geheugenruimte nodig heeft dan de klassieke formule van Kutta (1901) en ook voor een deel in extra precisie rekent. Blum (1957) heeft evenwel bewezen, dat dezelfde geheugenruimte-besparing kan worden bereikt voor de klassieke formule van Kutta. Huta (1956 & 1957) en Butcher (1963) geven formules van de orde 5 en 6. Butcher geeft tevens impliciete Runge-Kutta-formules analoog aan de Gauss-Legendre quadratuur-formules die bij q rechterlid-evaluaties een orde $2q$ halen. De procedure RK gepubliceerd door Naur (1960) gebruikt de klassieke 4^e orde-formule van Kutta met automatische stapvariatie gebaseerd op de discrepantie tussen integratie over 2 enkele stappen en integratie over één dubbele stap.

Zonneveld (1964) ontwierp formules van de orde 2 t/m 5 waarin, ten kosten van hoogstens één extra rechterlid-evaluatie, de *"discrepantie"*, dat is een benadering van de hoogste in rekening gebrachte Taylor-term, wordt verkregen. Met behulp hiervan wordt telkens de staplengte bepaald volgens een extrapolatie-formule die erop mikt de discrepantie 5% onder de tolerantie te houden.

Aangezien de genoemde Runge-Kutta-formules een behoorlijk hoge orde hebben (4 tot 6), kan men ermee een hoge precisie bereiken voor vrij grote

staplengte. De formules zijn evenwel minder efficient wanneer de (stabiele) Jacobiaan van het stelsel differentiaalvergelijkingen eigenwaarden met reëel deel veel kleiner dan 0 heeft, zoals bij stijve stelsels het geval is. De stabiliteitseis legt dan een veel sterkere beperking op aan de toelaatbare staplengte dan de nauwkeurigheidseis.

Stabiliteit

Het rekenen met computers maakte integratie over lange trajecten mogelijk, zodat de stabiliteit van numerieke integratie-formules van primair belang werd. Er zijn dan ook verscheidene onderzoekingen betreffende de stabiliteit verricht, met name voor lineaire meerstapsformules.

De algemene gedaante van een lineaire k-stapsformule is

$$y_{n+1} = \sum_{i=0}^{k-1} \alpha_i y_{n-i} + h \sum_{i=-1}^{k-1} \beta_i f(x_{n-i}, y_{n-i}), \tag{2.8}$$

waarbij α_i en β_i niet van f of y_i afhangen. Voor expliciete formules geldt $\beta_{-1} = 0$.

Als de basispunten x_{n-i}, i = -1, ..., k-1, equidistant zijn, dan zijn de coëfficiënten α_i en β_i constanten onafhankelijk van x_i.

Dahlquist (1956) toonde aan dat de orde van een stabiele lineaire k-stapsformule hoogstens k + 2 kan zijn. Hij introduceerde het begrip A-stabiliteit gedefinieerd als volgt:

Een lineaire k-stapsformule (2.8) is *A-stabiel* als, bij toepassing van de formule met vaste positieve staplengte h op een willekeurige differentiaalvergelijking van de vorm

$$\frac{d}{dx} y(x) = \lambda y(x), \tag{2.9}$$

waarbij λ een complexe constante met negatief reëel deel is, alle oplossingen van (2.8) voor toenemende n naar 0 convergeren.

Dahlquist (1963) bewees dat een expliciete lineaire k-stapsformule niet A-stabiel kan zijn en dat een impliciete A-stabiele formule van de orde hoogstens 2 is, waarbij de kleinste fout wordt bereikt voor de trapeziumregel.

Naast A-stabiele formules zijn van belang formules die voor een zo groot mogelijk gebied in de linkerhelft van het $h\lambda$-vlak stabiel zijn, welk gebied ook de oorsprong moet bevatten, aangezien 0 altijd een eigenwaarde van de Jacobiaan van (2.5) is en stabiliteit verzekerd moet zijn voor $h \downarrow 0$.

Verschillende auteurs ontwierpen predictor-corrector formules (indien de corrector slechts eenmaal of een vast eindig aantal malen wordt toegepast is de methode in wezen expliciet en dus niet A-stabiel) met een groter stabiliteitsgebied dan bestaande formules van dezelfde orde, o.a: Hull en Creemer (1963), Crane en Klopfenstein (1965) en Krogh (1966).

Verscheidene auteurs beschouwen een combinatie van een predictor en een corrector van dezelfde orde, ofschoon het uit nauwkeurigheidsoverwegingen meer voor de hand ligt een corrector met een predictor van een orde lager te combineren, vgl. Spijker (1968), p.48.*) Het blijkt dat dergelijke combinaties meestal een groter stabiliteitsgebied leveren.

Belangrijke bijzondere gevallen van (2.8) zijn de formules verkregen door numerieke integratie van het rechterlid, n.l. expliciete formules van Bashforth en Adams (1883) en impliciete van Moulton (1926), en de formules verkregen door numerieke differentiatie van het linkerlid, zie Curtiss en Hirschfelder (1952),

De impliciete formules verkregen door numerieke differentiatie zijn volgens Henrici (1962) uit stabiliteitsoverwegingen slechts bruikbaar voor orde hoogstens 6, en zelfs in deze gevallen minder nauwkeurig dan de formules van Adams-Moulton van dezelfde orde. Henrici verwerpt daarom die formules en behandelt ze alleen omdat ze nog nuttig kunnen zijn voor andere doeleinden dan stap-voor-stap integratie van differentiaalvergelijkingen. Curtiss en Hirschfelder (1952) en Gear (1968) gebruiken evenwel deze formules (van order hoogstens 6) voor het oplossen van stijve differentiaalvergelijkingen omdat ze interessante stabiliteitseigenschappen bezitten. Gear toont aan dat ze *stiffly stable* zijn, d.w.z. stabiel in een bepaalde rechthoek om de oorsprong en het aansluitende halfvlak links van een lijn parallel aan de imaginaire as. Voor het oplossen van de impliciete stelsels gebruikt Gear de iteratie-formule van Newton, waarvoor uiteraard berekening van de Jacobiaan noodzakelijk is. Bovendien bouwt hij in zijn programma automatische variatie van staplengte en orde in.

---

*) Zie ook Krogh (1966), die dit wel opmerkt op p.379, maar (ten onrechte) stelt dat de ordes gelijk moeten zijn om de foutterm te vinden.

Exponentiële aanpassing

Zoals reeds is opgemerkt heeft de Jacobiaan van een stelsel stijve differentiaalvergelijkingen een of meer eigenwaarden met reëel deel veel kleiner dan 0. Het stelsel in autonome vorm (2.4) t/m (2.6) heeft vanzelfsprekend ook een eigenwaarde 0, vgl. (2.7). M.a.w. stelsels stijve differentiaalvergelijkingen hebben een breed spectrum. Tussen een eigenwaarde met kleinste reële deel en 0 kunnen de andere eigenwaarden min of meer homogeen verspreid liggen ofwel de eigenwaarden zijn geconcentreerd in een klein aantal clusters (*cluster-spectrum*).

De bovengenoemde methoden van Curtiss-Hirschfelder en Gear zijn geschikt indien de eigenwaarden verspreid liggen. We zullen nu enige *cluster-methoden* noemen speciaal geschikt voor cluster-spectra. Deze methoden zijn gebaseerd op *exponentiële aanpassing*, waarvan het principe geïntroduceerd werd door Pope (1963).

Voor niet te kleine h wordt de oplossing benaderd door de exponentiële oplossing van het gelineariseerde probleem. Voor het stelsel (2.1) luidt de door Pope voorgestelde formule

$$y_{n+1} = y_n + hf + \{e^{hJ} - I - hJ\}\, J^{-1}(J^{-1}f_x + f), \tag{2.10}$$

waarbij f, f_x en J worden geëvalueerd in (x_n, y_n). Deze formule is van de orde 2. De factor tussen accoladen wordt in een Taylorreeks ontwikkeld zodat we krijgen

$$y_{n+1} = y_n + hf + h^2(f_x + Jf) \sum_{k=0}^{\infty} \frac{h^k J^k}{(k+2)!}\,. \tag{2.11}$$

Pope merkt op dat het sommeren van 20 à 30 termen in (2.11) goedkoper is dan het berekenen van (2.10) en dat het gebruik van deze formule de moeite loont omdat h groot gekozen kan worden. Dit laatste is voor formule (2.11) discutabel vanwege de slechte convergentie van deze reeks voor grote h.

Lawson (1967) stelt voor het stelsel (2.1) te transformeren in een stelsel dat als oplossing heeft de functie z gedefinieerd door

$$z(x) = \exp(-Ax)\, y(x), \tag{2.12}$$

waarbij A een benadering van de Jacobiaan J is, en daarna het getransformeerde stelsel met een Runge-Kutta-formule op te lossen. Voor het bepalen van A kan J eens, enkele malen of in elke stap opnieuw berekend worden. Het effect van stijfheid van het gegeven stelsel kan men aldus elimineren. Het ziet er echter naar uit dat men soms de moeilijkheden terug krijgt in de vorm van numerieke instabiliteit in de berekening van elke stap afzonderlijk.

Fowler & Warten (1967) brengen een expliciete 2-stapsformule met een extra rechterlid-evaluatie in een tussenpunt (zoals bij Runge-Kutta) en met een bepaalde exponentiële aanpassing. De formule is bedoeld voor het geval dat er een "grote spreiding van tijdconstanten is, waarvan de kleinste reëel is", hetgeen wil zeggen dat de eigenwaarden in twee reële ver uiteenliggende clusters liggen. Zij geven o.a. het volgende voorbeeld.

$$\frac{d}{dx}\, y(x) = \begin{pmatrix} -500.5 & 499.5 \\ 499.5 & -500.5 \end{pmatrix} y(x) + \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \qquad y(0) = \begin{pmatrix} -0.1 \\ +0.1 \end{pmatrix}.$$

De eigenwaarden van de Jacobiaan zijn -1 en -1000 en de oplossing luidt

$$y^1(x) = 2(1-e^{-x}) - 0.1\, e^{-1000x},$$

$$y^2(x) = 2(1-e^{-x}) + 0.1\, e^{-1000x}.$$

Liniger en Willoughby (1970) geven 3 impliciete eenstapsformules met exponentiële aanpassing voor spectra bestaande uit 2 of 3 clusters. De exponentiële aanpassing houdt in dat voor het centrum λ van elke cluster de oplossing in het punt x_{n+1} (m.a.w. voor de gebruikte staplengte h) exact overeenstemt met de oplossing van de lineaire vergelijking (2.9). Eén cluster-centrum ligt altijd in de oorsprong (hiervoor betekent de exponentiële aanpassing dus consistentie van de formule), het andere centrum is negatief reëel respectievelijk de andere twee centra zijn negatief reëel of toegevoegd complex.

De eerste formule luidt

$$y_{n+1} = y_n + h[\mu f_n + (1-\mu) f_{n+1}], \tag{2.13}$$

waarbij $f_n = f(x_n,y_n)$ en f_{n+1} analoog.
Deze formule is van de orde 1 (of voor het speciale geval $\mu = \frac{1}{2}$, de trapeziumregel, van de orde 2), en bovendien A-stabiel voor $\mu \leq \frac{1}{2}$. Voor exponentiële aanpassing in het centrum λ van een (ver van de oorsprong liggende) cluster moet gekozen worden:

$$(2.14) \qquad \mu = q^{-1} - (e^q-1)^{-1},$$

waarbij $q = -h\lambda$. Deze waarde van μ ligt in het interval $[0,\frac{1}{2}]$.

De tweede formule luidt

$$(2.15) \qquad y_{n+1} = y_n + \frac{h}{2}[(1-a)f_n+(1+a)f_{n+1}] + \frac{h^2}{4}[(b-a)J_nf_n-(b+a)J_{n+1}f_{n+1}],$$

waarbij $J_n = J(x_n,y_n)$.
Deze formule is van de orde 2 (of 3 als $b = 1/3$) en is A-stabiel in een bepaalde rechthoek van het (a,b)-vlak. Bovendien kunnen a en b in dit gebied zo gekozen worden dat de formule exponentieel wordt aangepast in 2 (ver van de oorsprong liggende) clusters met de restrictie, dat een toegevoegd complex paar niet te dicht bij de imaginaire as mag liggen. Hierbij zijn a en b functies van $q_i = -\lambda_i h$, waarbij λ_i (i=1,2) de centra der 2 clusters zijn.

De derde formule ontstaat uit de vorige door $b = 1/3$ te kiezen. Deze formule is van de orde 3 (of 4 als $a = 0$) en kan slechts in één ver van de oorsprong liggende cluster exponentieel aangepast worden.

Voor het oplossen van de stelsels vergelijkingen van deze impliciete methoden gebruiken Liniger en Willoughby het iteratie-proces van Newton. Merkwaardig is, dat zij als start voor deze iteratie de vector y_n gebruiken. Voor stelsels stijve differentiaalvergelijkingen levert een expliciete (predictor-) formule vaak geen duidelijk betere startwaarde.

Expliciete eenstapsformules (Taylor en Runge-Kutta) met exponentiële aanpassing zijn ontworpen door Van der Houwen (1970 & 1971). Hierin wordt een polynoombenadering gekozen, die tot een zekere orde met de Taylor-ontwikkeling overeenstemt. De hogere-graads coëfficiënten worden zo bepaald dat exponentiële aanpassing in 2 of 3 clusters wordt bereikt.

Aansluitend op deze gemodificeerde Taylor-benaderingen zijn Runge-Kutta-formules ontworpen, waarbij de parameters zo gekozen zijn dat slechts weinig geheugenruimte nodig is. Dit is vooral van belang bij het oplossen van zeer grote stelsels.

Component-afhankelijke parameters

Saul'yev (1964) introduceerde de *hopscotch* (d.i. "hinkspel")-methode voor het oplossen van de diffusie-vergelijking. Op elk tijdsniveau worden eerst de punten met even index berekend met de eenvoudigste expliciete formule, en daarna de punten met oneven index met een impliciete formule. Dit is nauwelijks bewerkelijker dan het expliciete schema en de staplengte kan 2 x zo groot gekozen worden. Gourlay (1969) breidde dit principe uit tot een grotere klasse van parabolische vergelijkingen.

Het idee blijkt ook bij het oplossen van stelsels stijve differentiaalvergelijkingen winst op te leveren. Toepassen van dit principe houdt dan in dat de vergelijkingen uit een gegeven stelsel op verschillende wijze behandeld worden. We zullen dit aan de hand van een voorbeeld illustreren. Beschouw het lineaire stelsel differentiaalvergelijkingen

$$(2.16) \qquad \frac{d}{dx}\, y(x) = Jy(x) + g(x),$$

waarbij $J = J(x)$ niet van y afhangt.
Schrijven we $J = J_1 + J_2$ zó dat $I - hJ_1$ gemakkelijk inverteerbaar is, dan vinden we hierbij de volgende formule

$$(2.17) \qquad y_{n+1} = [I+h(I-hJ_1)^{-1}J]y_n + h[I-hJ_1]^{-1}f_n$$

welke exact is als J en g constant zijn.
Stel in het bijzonder

$$J = \begin{pmatrix} -a & b \\ b & -a \end{pmatrix}, \quad J_1 = \begin{pmatrix} 0 & 0 \\ b & -a \end{pmatrix}$$

met $a > b > 0$ (a stijf als $a \approx b$).
Dit levert als stabiliteitsvoorwaarde

$$h \leq 2/\sqrt{a^2-b^2} = 2/\sqrt{\lambda_1\lambda_2},$$

terwijl voor de formule van Euler de stabiliteitsvoorwaarde luidt

$$h \leq 2/\max(|\lambda_1|,|\lambda_2|).$$

Dit betekent dat een aanzienlijke winst verkregen wordt voor het geval dat

$$|\lambda_1| >> |\lambda_2|.$$

In de hopscotch-methode voor grotere stelsels zou men J_1 zó kunnen kiezen dat de componenten y^i_{n+1} uit stelsels met twee onbekenden opgelost kunnen worden. Ook zou men *partieel impliciete* formules kunnen opstellen, waarbij telkens een klein aantal (zeg k) componenten simultaan worden berekend door oplossen van een stelsel van de orde k. Mogelijk kunnen ook Runge-Kutta-formules met componentsgewijs variërende parameters worden ontworpen om betere stabiliteitseigenschappen te krijgen.

Niet-lineariteit

Tot nu toe hebben de stabiliteitsbeschouwingen zich vrijwel uitsluitend beperkt tot stelsels lineaire differentiaalvergelijkingen. Een belangrijke uitzondering wordt gemaakt door Gourlay (1970), die laat zien dat de trapeziumregel (d.i. de nauwkeurigste A-stabiele methode), een instabiel gedrag vertoont voor de vergelijking

$$(2.18) \qquad \frac{d}{dx} y(x) = \lambda(x)\, y(x), \quad \lambda(x) \leq 0,$$

als $|\lambda(x)|$ monotoon afneemt en $h > 4/|\lambda(x_n)-\lambda(x_{n+1})|$.
Dit is dus een ernstige beperking als $|\lambda(x)|$ groot is en snel afneemt.

Voorbeeld

$$\lambda(x) = -\alpha^2(\beta-x) \qquad \text{voor } 0 \leq x \leq \beta,$$

$$\lambda(x) = 0 \qquad \text{voor } x \geq \beta,$$

met $\alpha = 10$, $\beta = 100$, $y(0) = 1$.
De exacte oplossing hiervan luidt

$$y(x) = \exp(-\alpha^2 x(\beta - x/2)), \qquad 0 \leq x \leq \beta$$

$$\exp(-\alpha^2\beta^2/2) \quad , \qquad x \geq \beta.$$

Gourlay stelt de volgende formule voor

$$(2.19) \qquad y_{n+1} = y_n + hf((x_n+x_{n+1})/2,(y_n+y_{n+1})/2),$$

die niet alleen de goede eigenschappen van de trapeziumregel bezit, maar ook in het bovengenoemde geval stabiel is. Gourlay geeft een analoge formule voor het probleem van Goursat, zijnde een beginwaarde-probleem bij een bepaalde hyperbolische differentiaalvergelijking. Hij merkt op dat de gesignaleerde vorm van instabiliteit in de praktijk inderdaad voorkomt, en niet onwaarschijnlijk is bij "strongly stiff problems".

Het is duidelijk dat bij het ontwerpen van methoden voor het oplossen van stelsels stijve differentiaalvergelijkingen het effect van niet-lineariteit op de stabiliteit van de methoden mede in ogenschouw genomen dient te worden.

Literatuur

Bashforth, F. and Adams, J.C.
Theories of Capillary action.
Cambridge Univ. Press (1883).

Blum, E.K.
A modification of the Runge-Kutta fourth-order method.
Numerical Note NN80. Ramo-Wooldridge Comp. Los Angeles (1957).

Butcher, J.C.
On the integration processes of A. Huta.
J. Australian Math. Soc. 3 (1963) 202.

Crane, R.L., and Klopfenstein, R.W.
A predictor-corrector algorithm with an increased range of absolute stability.
J. ACM 12 (1965) 227.

Curtiss, C.F. and Hirschfelder, J.O.
Integration of stiff equations.
Proc. Nat. Acad. Sci. U.S. 38 (1952) 235.

Dahlquist, G.
Convergence and stability in the numerical integration of ordinary differential equations.
Math. Scand. 4 (1956) 33.

Dahlquist, G.
A special stability problem for linear multistep methods.
BIT 3 (1963) 27.

Fowler, M.E. and Warten, R.M.
A numerical integration technique for ordinary differential equations with widely separated eigenvalues.
IBM J. Res. Develop. 11 (1967) 537.

Gear, C.W.
The automatic integration of stiff ordinary differential equations.
Proc. IFIP Congr. (1968) 187.

Gill, S.
A process for the step-by-step integration of differential equations in an automatic digital computing machine.
Proc. Cambridge Phil. Soc. 47 (1951) 96.

Gourlay, A.R.
The numerical solution of evolutionary partial differential equations.
in: Conf. on the Num. Sol. of Diff. Equations.
Lecture notes in mathematics, Springer (1969).

Gourlay, A.R.
A note on the trapezoidal methods for the solution of initial value problems.
Math. Comp. 24 (1970) 629.

Henrici, P.
Discrete variable methods in ordinary differential equations.
John Wiley and Sons, New York (1962).

Houwen, P.J. van der
One-step methods for linear initial value problems.
TW reports 122/70, 123/71, 130/71. Mathematisch Centrum, Amsterdam.

Hull, T.E. and Creemer, A.L.
Efficiency of predictor-corrector procedures.
J. ACM 10 (1963) 291.

Huta, A.
Une amélioration de la méthode de Runge-Kutta-Nyström pour la résolution numérique des équations differentielles du premier ordre.
Acta Fac. Nat. Univ. Comenian. Math. 1 (1956) 201.

Huta, A.

Contribution à la formule de sixième ordre dans la méthode de Runge-Kutta-Nyström.
Acta. Fac. Nat. Univ. Comenian. Math. 2 (1957) 21.

Krogh, F.T.

Predictor corrector methods of high order with improved stability characteristics.
J. ACM 13 (1966) 374.

Kutta, W.

Beitrag zur näherungsweisen Integration totaler Differentialgleichungen.
Z. Math. Phys. 46 (1901) 435.

Lawson, J.D.

Generalized Runge-Kutta processes for stable systems with large Lipschitz constants.
SIAM J. Num. Anal. 4 (1967) 372.

Liniger, W. and Willoughby, R.A.

Efficient integration methods for stiff systems of ordinary differential equations.
SIAM J. Num. Anal. 7 (1970) 47.

Moulton, F.R.

New methods in exterior ballistics.
Univ. Chicago Press (1926).

Naur, P. (ed.)

Report on the algorithmic language ALGOL 60.
Regnecentralen, Kopenhagen (1960).

Nordsieck, A.

On numerical integration of ordinary differential equations.
Math. Comp. 16 (1962) 22.

Pope, D.A.
An exponential method of numerical integration of ordinary differential equations.
C. ACM 6 (1963) 491.

Saul'yev, V.K.
Integration of equations of parabolic type by methods of nets.
Pergamon Press (1964).

Spijker, M.N.
Stability and convergence of finite-difference methods.
Proefschrift Leiden (1968).

Zonneveld, J.A.
Automatic numerical integration.
Math. Centre Tracts 8. Mathematisch Centrum, Amsterdam (1964).

3. Eenstapsmethoden

In dit hoofdstuk wordt nader ingegaan op integratiemethoden van de vorm

$$y_{n+1} = E(y_n), \tag{3.1}$$

waarin E een in het algemeen niet-lineaire operator is. Er zullen drie klassen van eenstapsmethoden beschouwd worden:

(1) methoden gebaseerd op herhaalde differentiatie van de differentiaalvergelijking (*Taylor-methoden*);
(2) methoden gebaseerd op herhaalde evaluatie van het rechterlid van de differentiaalvergelijking (*Runge-Kutta-methoden*);
(3) methoden gebaseerd op herhaalde evaluatie van het rechterlid en de Jacobiaan van de differentiaalvergelijking (*semi-Runge-Kutta-methoden*);

Voor deze drie typen van integratiemethoden zullen de begrippen *consistentie*, *convergentie* en *stabiliteit* uitvoerig besproken worden. In het bijzonder zal aandacht besteed worden aan de op het Mathematisch Centrum ontwikkelde integratieprocessen.

Integratieformules gebaseerd op herhaalde differentiatie

Zij gegeven de vectordifferentiaalvergelijking

$$\frac{dy}{dx} = f(y) \tag{3.2}$$

met de beginvoorwaarde

$$y(x_0) = y_0 \tag{3.3}$$

en laat $\{x_n\}$ een verzameling punten op de x-as zijn, gedefinieerd door de relatie

$$x_n = x_{n-1} + h_{n-1}, \qquad n = 1,2,\ldots, \tag{3.4}$$

waarin h_{n-1} de n-de integratiestap is.
Wanneer f(y) een voldoend aantal malen naar y differentieerbaar is, dan kan

men de volgende algemene eenstapsmethode definiëren:

$$\sum_{j=0}^{m_1} \alpha_j h_n^j y_{n+1}^{(j)} = \sum_{j=0}^{m_2} \beta_j h_n^j y_n^{(j)}, \qquad \alpha_0 = 1,$$

$$y_n^{(j)} = \frac{d^{j-1}}{dx^{j-1}} f(y(x))\Big|_{y=y_n}, \qquad y_n^{(0)} = y_n, \tag{3.5}$$

$$j = 1,2,\ldots,\ n = 0,1,2,\ldots\ .$$

Hierin stelt y(x) de oplossing van (3.2) door het punt (x_n,y_n) voor (*locaal analytische oplossing*).
De parameters α_j en β_j worden bepaald op grond van consistentie-, convergentie- en stabiliteitsvoorwaarden.

Wanneer $\alpha_j = 0$ voor $j = 1,2,\ldots,m_1$, dan kan y_{n+1} zonder meer uit het rechterlid van (3.5) berekend worden. De methode wordt *expliciet* genoemd. In alle andere gevallen ontstaat een algebraïsche vergelijking voor (de vector) y_{n+1}. De methode wordt dan *impliciet* genoemd.

Voorbeelden

Euler-formule

$$y_{n+1} = y_n + h_n y_n'. \tag{3.6}$$

Backward-Euler-formule

$$y_{n+1} - h_n y_{n+1}' = y_n. \tag{3.7}$$

Trapezium-regel

$$y_{n+1} - \tfrac{1}{2}h_n y_{n+1}' = y_n + \tfrac{1}{2}h_n y_n'. \tag{3.8}$$

$F^{(3)}$-formule van Liniger en Willoughby (1970)

$$(3.9)\qquad y_{n+1} - \tfrac{1}{2}(1+\alpha)\, h_n\, y'_{n+1} + \frac{1}{12}(1+3\alpha)\, h_n^2\, y''_{n+1} = y_n + \tfrac{1}{2}(1-\alpha)\, h_n\, y'_n + \frac{1}{12}(1-3\alpha)\, h_n^2\, y''_n.$$

In deze formule is α een parameter welke op grond van stabiliteitsoverwegingen bepaald wordt. Hierop zullen we later in dit hoofdstuk nog terugkomen.

Uit deze voorbeelden blijkt dat er sprake is van een gewogen Taylorontwikkeling van de numerieke oplossing in de punten $x = x_n$ en $x = x_{n+1}$. In het vervolg zullen we formules van het type (3.5) dan ook Taylor-achtige of kortweg Taylor-methoden noemen.

Integratieformules gebaseerd op herhaalde functie-evaluatie

In vele gevallen waarin de differentiatie van het rechterlid van de differentiaalvergelijking op moeilijkheden stuit, komt de volgende methode in aanmerking:

$$(3.10)\qquad \begin{aligned} y_{n+1} &= y_n + \sum_{j=0}^{m-1} \theta_j\, k_n^{(j)}, \\ k_n^{(j)} &= h_n\, f\Big(y_n + \sum_{l=0}^{m-1} \lambda_{jl} k_n^{(l)}\Big). \end{aligned}$$

Dit schema wordt een m-punts Runge-Kutta-schema genoemd. De Runge-Kutta-parameters θ_j en λ_{jl} volgen weer uit consistentie-, convergentie- en stabiliteitsvoorwaarden.

In het vervolg zullen we (3.10) karakteriseren door het array

$$(3.10')\qquad \begin{pmatrix} \Lambda \\ \theta \end{pmatrix},$$

waarin θ de rijvector $(\theta_0,\ldots,\theta_{m-1})$ en Λ de (m*m) matrix (λ_{jl}) voorstelt.

De Runge-Kutta-formule (3.10) wordt *expliciet* genoemd wanneer Λ een onderdriehoeksmatrix is en *impliciet* in de overige gevallen.

Voorbeelden

Heun-formule

$$(3.11) \qquad \begin{pmatrix} A \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Zie Heun (1900).

De volgende vier integratieformules ((3.12)-(3.14)) werden op het Mathematisch Centrum ontwikkeld (van der Houwen (1968, 1970a, 1971a)).

2-puntsgestabiliseerde Euler-formule

$$(3.12) \qquad \begin{pmatrix} A \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

4-punts gestabiliseerde Euler-formule

$$(3.13) \qquad \begin{pmatrix} A \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/64 & 0 & 0 & 0 \\ 0 & 1/20 & 0 & 0 \\ 0 & 0 & 5/32 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

3-punts gestabiliseerde Heun-formules

$$(3.14) \qquad \begin{pmatrix} A \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1/8 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} A \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Standaard Runge-Kutta-formule

$$(3.15) \qquad \begin{pmatrix} A \\ \theta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{pmatrix}.$$

Zie Runge (1895) en Kutta (1901).

Impliciete Gauss-formule

$$(3.16)\qquad \begin{pmatrix} \Lambda \\ \theta \end{pmatrix} = \begin{pmatrix} 1/4 & (3-2\sqrt{3})/12 \\ (3+2\sqrt{3})/12 & 1/4 \\ 1/2 & 1/2 \end{pmatrix}.$$

Zie Butcher (1964 a,b).

Integratieformules gebaseerd op de Jacobiaan

Het derde type integratieformule dat in deze bijdrage besproken zal worden is van de vorm

$$(3.17)\qquad \begin{aligned} y_{n+1} &= y_n + \sum_{j=0}^{m-1} \theta_j \, k_n^{(j)}, \\ k_n^{(j)} &= h_n \, R^{(j)}(h_n J_n) \, f(y_n + \sum_{l=0}^{j-1} \lambda_{jl} k_n^{(l)}), \end{aligned}$$

waarin J_n de Jacobiaan in $y = y_n$ voorstelt en $R^{(j)}(z)$ een rationale functie of polynoom is waarvan de vorm van j afhangt.

Formules van het type (3.7) noemt men *semi-impliciet* wanneer minstens één $R^{(j)}(z)$ rationaal is, aangezien dan de operator $R^{(j)}(h_n J_n)$ vermenigvuldiging met de inverse van een matrix met zich meebrengt. Wanneer alle $R^{(j)}(z)$ polynomen zijn, is de methode uiteraard *expliciet*.

We zullen in het vervolg methoden van het type (3.17) semi-Runge-Kutta-methoden noemen.

Voorbeelden

Rosenbrock-formule

$$(3.18)\qquad \begin{aligned} y_{n+1} &= y_n + k_n^{(1)}, \\ k_n^{(0)} &= h_n (I-(1-\tfrac{1}{2}\sqrt{2})h_n J_n)^{-1} \, f(y_n), \\ k_n^{(1)} &= h_n (I-(1-\tfrac{1}{2}\sqrt{2})h_n J_n)^{-1} \, f(y_n+(\tfrac{1}{2}\sqrt{2}-1)k_n^{(0)}). \end{aligned}$$

Zie Rosenbrock (1965).

Calahan-formule

$$y_{n+1} = y_n + \frac{3}{4} k_n^{(0)} + \frac{1}{4} k_n^{(1)},$$

(3.19)
$$k_n^{(0)} = h_n(I-\tfrac{1}{2}(1+\tfrac{1}{3}\sqrt{3})h_n J_n)^{-1} f(y_n),$$

$$k_n^{(1)} = h_n(I-\tfrac{1}{2}(1+\tfrac{1}{3}\sqrt{3})h_n J_n)^{-1} f(y_n - \tfrac{2}{3}\sqrt{3}\, k_n^{(0)}).$$

Zie Calahan (1968).

Op het Mathematisch Centrum zijn onlangs onderzocht de semi-impliciete formules

(3.20)
$$y_{n+1} = y_n + h_n(I - \frac{1}{2} h_n J_n + \frac{1}{12} h_n^2 J_n^2)^{-1} f(y_n)$$

en

$$y_{n+1} = y_n - 2k_n^{(0)} + 3k_n^{(1)},$$

(3.21)
$$k_n^{(0)} = h_n(I - \frac{1}{2} h_n J_n + \frac{1}{12} h_n^2 J_n^2)^{-1} (I - \frac{1}{4} h_n J_n) f(y_n),$$

$$k_n^{(1)} = h_n f(y_n + \frac{1}{3} k_n^{(0)}),$$

en de expliciete formule

$$y_{n+1} = y_n + \frac{3\alpha^2-1}{3\alpha^2} k_n^{(0)} + \frac{1}{3\alpha^2} k_n^{(1)},$$

(3.22)
$$k_n^{(0)} = h_n(1+\tfrac{1}{2}\alpha \frac{3\alpha-2}{3\alpha^2-1} h_n J_n + \tfrac{1}{2}\alpha^2 \frac{3\alpha^2-3\alpha+1}{(3\alpha^2-1)^2} h_n^2 J_n^2) f(y_n),$$

$$k_n^{(1)} = h_n f(y_n + \alpha k_n^{(0)}).$$

In formule (3.22) is α een nog onbepaalde parameter, waarmee de stabiliteit van de methode beïnvloed kan worden.

Consistentievoorwaarden

De eenstapsmethode (3.1) wordt een in $x = x_n$ *consistente* benadering van differentiaalvergelijking (3.2) genoemd wanneer voor elke oplossing $y(x)$ van (3.2) geldt

$$(3.23) \qquad y(x_{n+1}) - E(y(x_n)) \to 0 \qquad \text{als } h_n \to 0.$$

Consistentie betekent dus dat voor afnemende staplengte de *integratieformule* steeds meer op de differentiaalvergelijking gaat lijken. Alhoewel dit niet garandeert dat de *oplossingen* van (3.1) en (3.2) voor $h_n \to 0$ steeds meer op elkaar gaan lijken, dus *convergentie*, is (3.23) toch een voorwaarde die men in elk geval zal willen stellen.

De eenstapsformule (3.1) wordt consistent van de orde p genoemd wanneer

$$(3.24) \qquad y(x_{n+1}) - E(y(x_n)) = O(h_n^{p+1}) \quad \text{als } h_n \to 0.$$

Men bepaalt in het algemeen de orde van consistentie door ontwikkeling van $y(x_{n+1})$ en $E(y_n)$ in machten van h_n.

Consistentie van Taylor-methoden

Substitutie van een oplossing $y(x)$ in (3.5) en ontwikkeling in Taylorreeksen geeft voor $h_n \to 0$

$$y(x_n+h_n) - E(y(x_n)) = \sum_{j=0}^{m_1} \alpha_j h_n^j y^{(j)}(x_n+h_n) - \sum_{j=0}^{m_2} \beta_j h_n^j y^{(j)}(x_n) =$$

$$= (1-\beta_0)y(x_n) + (1-\alpha_1+\beta_1)h_n y'(x_n) +$$

$$+ \tfrac{1}{2}(1+2\alpha_1+2\alpha_2-2\beta_2)h_n^2 y''(x_n) +$$

$$+ \tfrac{1}{6}(1+3\alpha_1+6\alpha_2+6\alpha_3-6\beta_3)h_n^3 y'''(x_n) + O(h_n^4).$$

In tabel 3.1 zijn de consistentievoorwaarden voor p = 1,2,3 en 4 weergegeven.

Tabel 3.1. Consistentievoorwaarden voor formule (3.5)

p	β_0	β_1	β_2	β_3	β_4
1	1	$1 + \alpha_1$			
2	1	$1 + \alpha_1$	$\frac{1}{2} + \alpha_1 + \alpha_2$		
3	1	$1 + \alpha_1$	$\frac{1}{2} + \alpha_1 + \alpha_2$	$\frac{1}{6} + \frac{1}{2}\alpha_1 + \alpha_2 + \alpha_3$	
4	1	$1 + \alpha_1$	$\frac{1}{2} + \alpha_1 + \alpha_2$	$\frac{1}{6} + \frac{1}{2}\alpha_1 + \alpha_2 + \alpha_3$	$\frac{1}{24} + \frac{1}{6}\alpha_1 + \frac{1}{2}\alpha_2 + \alpha_3 + \alpha_4$

Toepassingen

Toepassing van tabel 3.1 op de formules (3.6) -(3.9) levert het volgende overzicht:

Euler	: p = 1
Backward - Euler	: p = 1
Trapezium-regel	: p = 2
$F^{(3)}$-formule van Liniger-Willoughby	: p = 3

Consistentie van Runge-Kuttamethoden

Voor de Runge-Kutta-methoden zijn de consistentievoorwaarden aanzienlijk gecompliceerder dan voor de Taylor-achtige methoden. We zullen ons hier dan ook beperken tot expliciete Runge-Kutta-formules en slechts de eerste 3 termen van de Taylor-ontwikkeling van (3.23) in het punt x_n geven: Uit (3.10) volgt dat voor $h_n \to 0$ geldt

$$y(x_n+h_n) - E(y(x_n)) = (1-\beta_1)h_n y'(x_n) + (\tfrac{1}{2} -\beta_2)h_n^2 y''(x_n) +$$

$$+ h_n^3[(\tfrac{1}{6} -\beta_3)J(y(x_n))y''(x_n) + \tfrac{1}{2}(\tfrac{1}{3} -\beta_{31})(y'''(x_n) - J(y(x_n))y''(x_n)] +$$

$$+ O(h_n^4),$$

waarin

$$\beta_1 = \sum_{j=0}^{m-1} \theta_j,$$

$$\beta_2 = \sum_{j=1}^{m-1} \theta_j \sum_{l=0}^{j-1} \lambda_{jl},$$

$$\beta_3 = \sum_{j=2}^{m-1} \theta_j \sum_{l=0}^{j-1} \lambda_{jl} \sum_{i=0}^{l-1} \lambda_{li},$$

$$\beta_{31} = \sum_{j=1}^{m-1} \theta_j \left[\sum_{l=0}^{j-1} \lambda_{jl}\right]^2.$$

In tabel 3.2 zijn de consistentievoorwaarden voor p = 1,2 en 3 gegeven.

Tabel 3.2. Consistentievoorwaarden voor formule (3.10)

p	β_1	β_2	β_3	β_{31}
1	1			
2	1	1/2		
3	1	1/2	1/6	1/3

Toepassingen

Met behulp van deze tabel kan eenvoudig de orde van consistentie van de voorbeelden (3.11) - (3.15) bepaald worden. We vinden het volgende overzicht:

Heun : $p = 2$
2-punts gestabiliseerde Euler: $p = 1$
4-punts gestabiliseerde Euler: $p = 1$
3-punts gestabiliseerde Heun : $p = 2$
Standaard Runge-Kutta : $p \geq 3$

De Runge-Kutta-formule (3.15) blijkt bij een meer volledige analyse dan hier gegeven is, een orde van consistentie 4 te hebben evenals de impliciete Gauss-formule.

Consistentie van semi-Runge-Kutta-methoden

Evenals bij het consistentie-onderzoek van de Taylor- en Runge-Kutta-methoden stelt men voor de semi-Runge-Kutta-methoden weer een Taylorreeks op. We zullen dit illustreren aan de hand van formule (3.20) en verder zonder bewijs de orde van consistentie geven van de overige voorbeelden van semi-Runge-Kutta-formules.

Uit (3.20) volgt voor $h_n \to 0$

$$y(x_n+h_n) - E(y(x_n)) = y(x_n+h_n) - y(x_n) -$$

$$- h_n(I - \tfrac{1}{2} h_n J(y(x_n)) + \tfrac{1}{12} h_n^2 J^2(y(x_n)))^{-1} f(y(x_n)) =$$

$$= y(x_n+h_n) - y(x_n) -$$

$$- h_n(I + \tfrac{1}{2} h_n J(y(x_n)) - \tfrac{1}{12} h_n^2 J^2(y(x_n)) +$$

$$+ \tfrac{1}{4} h_n^2 J^2(y(x_n)))f(y(x_n)) + O(h_n^4) =$$

$$= y(x_n+h_n) - y(x_n) - h_n y'(x_n) - \tfrac{1}{2} h_n^2 y''(x_n) -$$

$$- \tfrac{1}{6} h_n^3 J^2(y(x_n))f(y(x_n)) + O(h_n^4) = O(h_n^3).$$

De orde van consistentie is dus 2.

Op analoge wijze kan men voor de andere semi-Runge-Kutta-methoden de orde van consistentie bepalen:

Rosenbrock	:	p = 2
Calahan	:	p = 3
Formule (3.20)	:	p = 2
Formule (3.21)	:	p = 3
Formule (3.22)	:	p = 3

Convergentievoorwaarden

Zoals al opgemerkt is mag uit de consistentie van een integratieformule niet zonder meer geconcludeerd worden dat de numerieke oplossing y_n voor $h_n \to 0$ naar de analytische oplossing convergeert. We zullen nu laten zien dat voor de drie in dit hoofdstuk gegeven typen integratieformules geldt dat consistentie van orde $p \geq 1$ convergentie impliceert.

Laat $\tilde{y}(x)$ de analytische oplossing van (3.2) met beginvoorwaarde (3.3) voorstellen en $y(x)$ de locaal analytische oplossing. We definiëren nu de *discrepantie* in het punt $x = x_n$ door

$$\rho_n = y(x_{n+1}) - y_{n+1} \tag{3.25}$$

en de *discretiseringsfout* door

$$\varepsilon_n = \tilde{y}(x_n) - y_n. \tag{3.26}$$

Aangezien zowel $\tilde{y}(x)$ als $y(x)$ aan de differentiaalvergelijking (3.2) voldoen geldt (middelwaardestelling)

$$\tilde{y}'(x) - y'(x) = f(\tilde{y}) - f(y) = J(\bar{y})(\tilde{y}(x)-y(x)), \tag{3.27}$$

waarin $\bar{y}(x)$ een functie is in de omgeving van $y(x)$ en $J(\bar{y})$ de Jacobiaan van de operater $f(y)$ in $y = \bar{y}$ is. Uit (3.27) volgt

$$\tilde{y}(x_{n+1}) - y(x_{n+1}) = \exp[\int_{x_n}^{x_{n+1}} J(\bar{y}(x))dx].\varepsilon_n = \exp(h_n J(\bar{\bar{y}}_n)).\varepsilon_n, \tag{3.28}$$

waarin $\bar{\bar{y}}_n$ een punt in de omgeving van y_n is.

Met behulp van (3.25) en (3.28) vinden we

$$\begin{aligned} \varepsilon_{n+1} &= \tilde{y}(x_{n+1}) - y(x_{n+1}) + y(x_{n+1}) - y_{n+1} \\ &= \exp(h_n J(\bar{\bar{y}}_n))\varepsilon_n + \rho_n. \end{aligned} \tag{3.29}$$

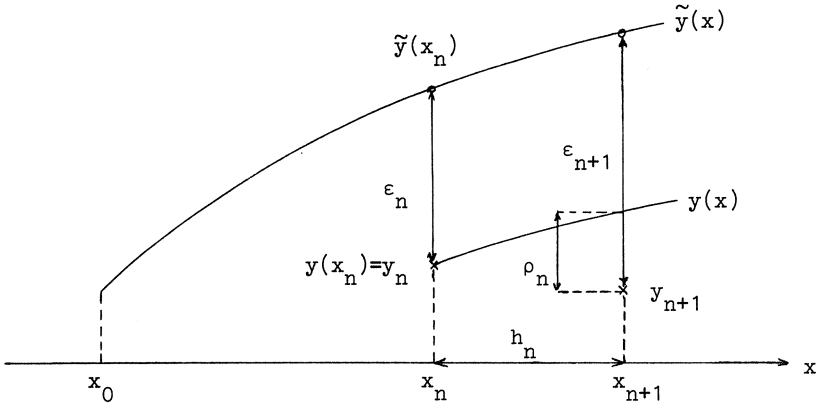


fig. 3.1. Illustratie van ρ_n en ε_n

Verder volgt uit de definitie van orde van consistentie dat voor de drie typen integratieformules geldt

$$\rho_n = O(h_n^{p+1}). \tag{3.30}$$

We definiëren nu

$$X = \sum_{n=0}^{N} h_n, \qquad h_{max} = \max_{0\le n\le N} h_n, \quad C = \max_{0\le n\le N} \frac{||\rho_n||}{h_n^{p+1}},$$

$$E_{max} = \max_{0\le n\le N} ||\exp(h_n J(\bar{\bar{y}}_n))||, \quad J_{max} = \max_{0\le n\le N} ||J(\bar{\bar{y}}_n)||. \tag{3.31}$$

waarin N+2 het aantal netpunten in het intergratie-interval $[x_0,x_0+X]$ is en $||\ ||$ de een of andere operator-norm voorstelt.

Stelling 3.1

Wanneer $E_{max} \le 1$ dan is

$$\max_{0\le n\le N} ||\varepsilon_{n+1}|| \le CXh_{max}^{p}.$$

Bewijs

Uit (3.24), (3.23) en (3.26) volgt

$$||\varepsilon_{n+1}|| \leq \sum_{j=0}^{n} Ch_j^{p+1} \leq Ch_{max}^{p} \sum_{j=0}^{n} h_j \leq CXh_{max}^{p}, \quad n=0,1,2,\ldots,N.$$

Stelling 3.2

Wanneer $E_{max} > 1$ dan is

$$\max_{0\leq n\leq N} ||\varepsilon_{n+1}|| < C \frac{\exp(XJ_{max}) - 1}{J_{max}} h_{max}^{p}.$$

Bewijs

Uit (3.29), (3.30) en (3.31) volgt

$$||\varepsilon_{n+1}|| \leq \exp(h_n J_{max}) ||\varepsilon_n|| + Ch_n^{p+1} \leq$$

$$\leq C[h_n^{p+1} + h_{n-1}^{p+1} \exp(h_n J_{max}) + h_{n-2}^{p+1} \exp(h_n+h_{n-1})J_{max} + \ldots$$

$$\ldots + h_0^{p+1} \exp(h_n+h_{n-1}+\ldots+h_1)J_{max}] \leq$$

$$\leq Ch_{max}^{p}[h_n + h_{n-1}\exp(h_n J_{max}) + \ldots + h_0 \exp(h_n+\ldots+h_1)J_{max}]$$

$$< Ch_{max}^{p} \int_0^X \exp(xJ_{max})dx = C \frac{(\exp(XJ_{max}) - 1}{J_{max}} h_{max}^{p}.$$

In deze bijdrage zullen we niet verder op het convergentie-probleem ingaan. Een meer gedetailleerde behandeling vindt men in Henrici (1962). Daarentegen zullen we onze aandacht van nu af aan op het stabiliteits-probleem concentreren aangezien dit het centrale punt is waar het in de numerieke analyse van stijve differentiaalvergelijkingen omdraait.

Stabiliteit

Wanneer men gaat rekenen met een minstens eerste orde convergente integratieformule weet men dat voor voldoende kleine integratiestappen en exacte rekenoperaties de gezochte oplossing benaderd wordt. Maar wat gebeurt er wanneer de rekenoperaties niet exact zijn en een rij eindige integratiestappen $\{h_n\}_0^N$ gegeven wordt? Deze vraagstelling leidt tot het begrip stabiliteit van een integratieformule.

Laat y_n^* de vector zijn die onstaat wanneer de vector y_n verstoord wordt bijvoorbeeld door afrondingsfouten. We definiëren nu de *numerieke fout* ε_n^* door

$$\varepsilon_n^* = y_n^* - y_n. \tag{3.32}$$

Wanneer we van nu af aan aannemen dat geen afrondingsfouten in de toepassing van (3.1) optreden, dan geldt

$$\varepsilon_{n+1}^* = y_{n+1}^* - y_{n+1} = E(y_n^*) - E(y_n) = E'(\bar{y}_n)\,\varepsilon_n^*, \tag{3.33}$$

waarin $\bar{y}_n$ een vector in de omgeving van y_n^* en y_n en $E'(\bar{y}_n)$ de afgeleide naar y van de operator in het punt $y = \bar{y}_n$ is.

Het ligt nu voor de hand formule (3.1) stabiel in het punt $(x_n,\bar{y}_n)$ te noemen wanneer

$$||E'(\bar{y}_n)|| < 1, \tag{3.34}$$

aangezien dan een eenmaal gemaakte verstoring direct weer gedempt wordt. De integratie van een gegeven beginwaarde-probleem met een bepaalde integratieformule zullen we een stabiel proces noemen wanneer de operator $E'(y)$ telkens in de omgeving van de numerieke oplossing y_n, $n = 0,1,\ldots,N$, een norm kleiner dan 1 heeft. We zullen een operator E waarvoor geldt $||E(y)|| < ||y||$ voor het gemak '*contraherend*' noemen. We merken op dat deze definitie van contraherend iets ruimer is dan de gebruikelijke: $||E(y) - E(y^*)|| < ||y - y^*||$.

Indien de operator lineair is, geldt $E'(y) = E' = E$, zodat instabiliteit in de zin van (3.34) betekent dat zowel de fouten als de numerieke oplossing exponentieel toenemen. Voor niet-lineaire operatoren kan het zijn

dat E contraherend is en E'(y) niet-contraherend in de omgeving van $\{y_n\}_{n=0}^{N}$ is. In dergelijke gevallen daalt de numerieke oplossing, alhoewel de afrondingsfouten kunnen toenemen. Instabiliteit is dan niet eenvoudig te detecteren op grond van de resultaten. Dit verraderlijke gedrag zullen we met een voorbeeld illustreren.

Voorbeeld

Beschouw eerst de lineaire vergelijking

$$y' = -y. \tag{3.35}$$

Passen we op deze vergelijking de Heun-formule (3.11) toe, dus

$$y_{n+1} = y_n + \frac{1}{2} h_n \left(f(y_n) + f(y_n + h_n f(y_n))\right), \tag{3.11'}$$

dan vinden we

$$\begin{aligned} E(y_n) &= (1 - h_n + \tfrac{1}{2} h_n^2)\, y_n, \\ E'(\bar{y}_n) &= 1 - h_n + \tfrac{1}{2} h_n^2. \end{aligned} \tag{3.36}$$

Hieruit volgt eenvoudig de stabiliteitsvoorwaarde

$$0 < h_n < 2. \tag{3.37}$$

Wanneer aan (3.37) voldaan is, zijn E en E' beide contraherend. Vervolgens beschouwen we de niet-lineaire vergelijking

$$y' = -y^2. \tag{3.38}$$

Toepassing van de Heun-formule (3.11) levert

$$\begin{aligned} E(y_n) &= (1 - h_n y_n + h_n^2 y_n^2 - \tfrac{1}{2} h_n^3 y_n^3)\, y_n, \\ E'(\bar{y}_n) &= 1 - 2h_n \bar{y}_n + 3h_n^2 \bar{y}_n^2 - 2h_n^3 \bar{y}_n^3. \end{aligned} \tag{3.39}$$

De stabiliteitsvoorwaarde (3.34) geeft voor dit voorbeeld in benadering

$$0 < h_n < \frac{1.3}{\bar{y}_n} . \tag{3.40}$$

De operator E is echter contraherend in alle punten (h_n, y) die voldoen aan

$$0 < h_n < \frac{2}{y}. \tag{3.41}$$

We merken op dat de contractie-voorwaarden in het lineaire geval niet en in het niet-lineaire geval wel van y afhangen.

Een tweede verraderlijk verschijnsel dat men bij niet-lineaire vergelijkingen kan treffen is dat het integratieproces in een punt terecht kan komen waar de operator E als de eenheidsoperator werkt, zodat de oplossing blijft "hangen".

Voorbeeld

Stel dat we de gestabiliseerde Euler-formule (3.12) toepassen op vergelijking (3.38). We vinden dat

$$\begin{aligned} E(y_n) &= (1 - h_n y_n + 2h_n^2 y_n^2 - h_n^3 y_n^3)\, y_n, \\ E'(y_n) &= 1 - 2h_n \bar{y}_n + 6h_n^{2-2} y_n^{2-2} - 4h_n^{3-3} y_n^{3-3}. \end{aligned} \tag{3.42}$$

In figuur 3.2 zijn $E(y_n)/y_n$ en $E'(y_n)$ als functie van $h_n y_n$ geschetst.

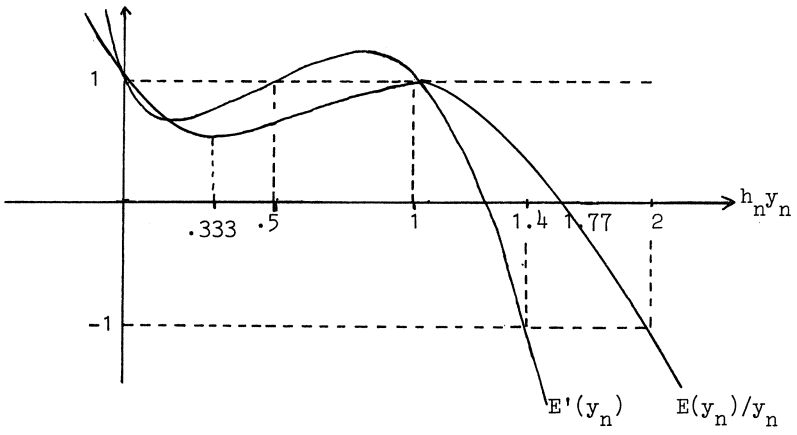


fig. 3.2. $E(y_n)/y_n$ en $E'(y_n)$ als functie van $h_n y_n$

Uit deze figuur lezen we af dat E contraherend is in de punten (h_n, y_n) die voldoen aan

$$(3.43) \qquad 0 < h_n < \frac{1}{y_n}, \frac{1}{y_n} < h_n < \frac{2}{y_n}$$

en dat $E'(y_n)$ stabiel is wanneer

$$(3.44) \qquad 0 < h_n < \frac{.5}{y_n}, \frac{1}{y_n} < h_n < \frac{1.4}{y_n} .$$

Voor we verdere conclusies uit figuur 3.2 zullen trekken, geven we de resultaten van een aantal experimenten uitgevoerd op de EL X8 van het Mathematisch Centrum. Deze experimenten werden gedaan om de stabiliteitseigenschappen van een integratiemethode toegepast op een niet-lineaire vergelijking te leren kennen, dus niet om de nauwkeurigheid van een integratiemethode te onderzoeken. In de berekeningen was h_n steeds .05 en werd elke waarde y_n verstoord met 1% waarbij het teken van de verstoring random gekozen werd. In tabel 3.3 zijn de y_n^*-waarden opgenomen die verkregen werden voor een aantal verschillende beginwaarden y_0.

We bespreken nu de verschillende kolommen van tabel 3.3.

$\underline{h_n y_0 = .5}$

Bij deze start aan de rand van het eerste stabiliteitsinterval wordt de analytische oplossing $(y(x) = y_0/(xy_0+1))$ althans enigszins benaderd.

$\underline{h_n y_0 = .8}$

De start vindt plaats van uit een instabiel gebied maar binnen het contractie-gebied van E. Weliswaar vindt er in de eerste 6 stappen foutenopbouw plaats, maar de numerieke oplossing neemt wel af. De benadering is uiteraard slecht.

Tabel 3.3 Resultaten van formule (3.12) toegepast op vergelijking (3.38) met $h_n = .05$ en $\varepsilon_n^* = y_n/100$

x_n	$h_n y_0 = .5$	$h_n y_0 = .8$	$h_n y_0 = 1$	$h_n y_0 = 1.4$	$h_n y_0 = 1.5$	$h_n y_0 = 1.8$
0	10.0	16.0	20.0	28.0	30.0	36.0
.05	8.75	15.5	20.0	21.7	18.8	- 5.47
.10	7.62	15.1	20.2	21.7	18.9	- 8.02
.15	6.44	14.2	20.0	21.4	18.6	-14.1
.20	5.43	13.2	19.8	21.1	18.3	-41.9
.30	4.02	11.0	19.8	20.1	18.1	-10^8
.40	3.14	8.74	20.2	21.2	18.2	$<-10^{100}$
.50	2.55	6.57	20.6	21.4	18.3	
.60	2.13	4.87	20.9	21.5	18.5	
.70	1.83	3.71	21.2	21.6	18.7	
.80	1.54	2.83	20.7	21.0	18.1	
.90	1.35	2.29	20.6	20.9	17.8	
1.00	1.20	1.91	20.6	20.8	17.4	
$\tilde{y}(1)$	$\frac{10}{11}$	$\frac{16}{17}$	$\frac{20}{21}$	$\frac{28}{29}$	$\frac{30}{31}$	$\frac{36}{37}$

$h_n y_0 = 1$

De numerieke oplossing blijft "hangen" omdat E niet contraherend is bij deze start.

$h_n y_0 = 1.4$

Een start aan de rand van het tweede stabiliteits interval en binnen het contractie-gebied van E. De numerieke oplossing daalt waarmee echter ook de contractie afneemt tot ongeveer 1. Dan is dezelfde situatie ontstaan als in het derde experiment:

$h_n y_0 = 1.5$

De contractie is bij deze start zo groot dat de drempel $h_n y_n = 1$ direct overschreden wordt. De numerieke oplossing zal blijven afnemen. Er vindt echter foutenopbouw plaats tot dat het eerste stabiliteitsgebied bereikt is.

$h_n y_0 = 1.8$

Weliswaar wordt gestart binnen het contractiegebied van E, maar $E(y_0)/y_0$ is nu negatief zodat direct een duidelijke instabiliteit optreedt.

Stabiliteitsfunctie en numerieke versterkingsfactoren

Stel dat we een eenstapsmethode toepassen op de modelvergelijking

(3.45) $\quad y' = \delta y,$

waarin δ een willekeurig complex getal is. We zullen dan voor alle in dit hoofdstuk beschreven integratie-formules een betrekking krijgen van de vorm

(3.46) $\quad y_{n+1} = R(h_n \delta)\, y_n,$

waarin R(z) een rationale functie of een polynoom is, welke uitsluitend

afhangt van de integratieformule en niet van δ, h_n of y_n. R(z) zullen we de *stabiliteitsfunctie* van de integratieformule en $R(h_n\delta)$ de *numerieke versterkingsfactor* of *karakteristieke wortel* behorend bij δ noemen.

In tabel 3.4 zijn de stabiliteitsfuncties van een aantal reeds ter sprake gekomen eenstapsformules opgenomen.

Tabel 3.4 Stabiliteitsfuncties

methode	R(z)
Taylor-methode (3.5)	$\dfrac{\sum_0^m \beta_j z^j}{\sum_0^m \alpha_1 z^j}$
3-punts expliciete Runge-Kutta-formule (3.10)	$1 + \beta_1 z + \beta_2 z^2 + \beta_3 z^3$
Heun-formule (3.11)	$1 + z + \frac{1}{2}z^2$
2-punts gestabiliseerde Euler-formule (3.12)	$1 + z + z^2$
4-punts gestabiliseerde Euler-formule (3.13)	$1 + z + \frac{5}{32} z^2 + \frac{1}{128} z^3 + \frac{1}{8192} z^4$
3-punts gestabiliseerde Heun-formules (3.14)	$1 + z + \frac{1}{2} z^2 + \frac{1}{16} z^3$ $1 + z + \frac{1}{2} z^2 + \frac{1}{4} z^3$
Standaard Runge-Kutta-formule (3.15)	$1 + z + \frac{1}{2} z^2 + \frac{1}{6} z^3 + \frac{1}{24} z^4$
Rosenbrock-formule (3.18)	$\dfrac{1 + (\sqrt{2}-1)z}{1 - (2-\sqrt{2})z + (\frac{3}{2}-\sqrt{2})z^2}$
Calahan-formule (3.19)	$\dfrac{1 - \frac{1}{3}\sqrt{3}z - \frac{1}{6}(1 + \sqrt{3})z^2}{1 - (1 + \frac{1}{3}\sqrt{3})z + \frac{1}{6}(2 + \sqrt{3})z^2}$
Formules (3.20) en (3.21)	$\dfrac{1 + \frac{1}{2} z + \frac{1}{12} z^2}{1 - \frac{1}{2} z + \frac{1}{12} z^2}$
Formule (3.22)	$1 + z + \frac{1}{2} z^2 + \frac{1}{6} z^3 + \frac{1}{6} \alpha \frac{3\alpha^2 - 3\alpha + 1}{(3\alpha^2 - 1)^2} z^4$

We merken op dat in het geval van de Taylor-methode de stabiliteitsfunctie deze methode volledig vastlegt. Dit is niet het geval met de Runge-Kutta- en semi-Runge-Kutta-methoden.

Het belang van de stabiliteitsfunctie ligt in de volgende stelling.

Stelling 3.3

Voor lineaire differentiaalvergelijkingen waarvan de Jacobiaan een normaal-matrix is, is de eenstapsmethode (3.1) stabiel in de zin van (3.34) dan en slechts dan wanneer de versterkingsfactoren behorend bij de eigenwaarden van de Jacobiaan binnen de eenheidscirkel liggen.

Bewijs

Past men een eenstapsmethode met stabiliteitsfunctie R(z) toe op de lineaire vergelijking

$$y' = Jy,$$

waarin J een matrix-operator is, dan geldt volgens (3.46)

$$y_{n+1} = R(h_n J_n)\, y_n.$$

De operator $R(h_n J_n)$ heeft dezelfde eigenfuncties als J_n en is dus ook een normaal-matrix. Wanneer we met $||\ ||$ de operator-norm behorend bij de Euclidische vector-norm aangeven, dan geldt

$$||E'(\bar{y}_n)|| = ||R(h_n J_n)|| = \max_{\lambda} |\, R_n(h_n \lambda)| < 1,$$

waarmee de stelling bewezen is.

In de praktijk stelt men ook in de gevallen waarin niet aan de voorwaarden van de stelling voldaan is, de eis dat de versterkingsfactoren behorend bij de eigenwaarden van de Jacobiaan J_n binnen de eenheidscirkel liggen. Dit komt overeen met het lineariseren van de differentiaalvergelijking in de roosterpunten x_n (vergelijk de in de inleiding gegeven stabiliteitsbeschouwing). Een dergelijk locaal stabiliteitscriterium moet

met de nodige voorzichtigheid gehanteerd worden zoals bleek uit het voorbeeld van Gourlay dat in hoofdstuk 2 besproken is. Er zijn echter ook gevallen waarin de integratieformule volgens (3.34) juist een groter stabiliteitsgebied heeft dan volgens het locale criterium.

Voorbeelden

Beschouw nog eens de Heun-formule (3.11) toegepast op vergelijking (3.38). De Jacobiaan in $x = x_n$ wordt gegeven door de scalar $-2y_n$ en de bijbehorende versterkingsfactor volgens tabel 3.4 door

$$R(-2h_n y_n) = 1 - 2h_n y_n + 2h_n^2 y_n^2 .$$

Deze versterkingsfactor ligt binnen de eenheidscirkel wanneer

$$(3.47) \qquad 0 < h_n < \frac{1}{y_n} .$$

We hebben echter gezien dat een strenge analyse oplevert (zie voorwaarde 3.40)

$$(3.40') \qquad 0 < h_n < \frac{1.3}{\bar{y}_n} \sim \frac{1.3}{y_n} .$$

Een tweede voorbeeld is de gestabiliseerde Euler-formule (3.12). Het locale stabiliteitscriterium volgt uit de stabiliteitsfunctie (zie tabel 3.4)

$$R(z) = 1 + z + z^2 ,$$

namelijk

$$(3.48) \qquad 0 < h_n < \frac{.5}{y_n} .$$

Volgens (3.44) levert een niet-locale analyse echter

$$0 < h_n < \frac{.5}{y_n} , \frac{1}{y_n} < h_n < \frac{1.4}{y_n} ,$$

waarin we weer $\bar{y}_n \sim y_n$ hebben verondersteld.

In dit hoofdstuk zullen we ons uitsluitend bezighouden met locale stabiliteitsanalyses, dus met de stabiliteitsfunctie R(z). Hiervoor definiëren we het *stabiliteitsgebied*

$$S : \{z \mid |R(z)| < 1\}. \tag{3.49}$$

Wanneer S het gehele linkerhalfvlak Re z < 0 bevat, spreekt men van *A-stabiliteit* (Dahlquist (1963), zie ook hoofdstuk 2, blz. 22). Verder zullen we de absolute waarde van het snijpunt van de randkromme ∂S van S met de negatieve en imaginaire as respectievelijk *reële* en *imaginaire stabiliteitsgrens* noemen; de straal van de cirkel die in het linker halfvlak nog juist tot $S \cup \partial S$ behoort, zullen we *absolute stabiliteitsgrens* noemen.

Consistente stabiliteitsfuncties

Voordat we overgaan tot de constructie van stabiliteitsfuncties met een voorgeschreven stabiliteitsgebied S, geven we de voorwaarden aan waaronder een stabiliteitsfunctie compatibel is met een p-de orde consistente integratieformule.

Stel dat de eenstapsformule (3.1) p-de orde consistent is en een stabiliteitsfunctie

$$R(z) = \frac{\sum_{j=0}^{m_2} \beta_j z^j}{\sum_{j=0}^{m_1} \alpha_j z^j} \tag{3.50}$$

heeft. Volgens de definitie van R(z) is (3.1), toegepast op de modelvergelijking (3.45), te schrijven als (3.46). Formule (3.46) is echter identiek aan de Taylor-methode (3.5) toegepast op de modelvergelijking. Hieruit volgt dat de coëfficienten α_j en β_j moeten voldoen aan de in tabel 3.1 gegeven voorwaarden. Functies van de vorm (3.50) die aan de voorwaarden genoemd in tabel 3.1 voldoen worden p-de orde consistente stabiliteitsfuncties genoemd. We merken hierbij op dat een p-de orde consistente stabiliteitsfunctie in het algemeen niet impliceert dat de bijbehorende integratieformule consistent van de orde p is. Dit is alleen het geval wanneer men zich beperkt tot de klasse van lineaire differentiaalvergelijkingen.

De functies R(z) die bij gegeven waarden van m_1 en m_2 een maximale orde van consistentie hebben, worden *Padé-benaderingen* van exp z genoemd (wij zullen in dit hoofdstuk de niet-rationale Padé-benaderingen, dus $m_1 = 0$, *Padé-polynomen* noemen). Zij zijn eenduidig bepaald en hebben een orde van consistentie $p = m_1 + m_2$. In tabel 3.5 zijn een aantal Padé-benaderingen opgenomen.

Tabel 3.5 Padé-benaderingen van exp z

	$m_2 = 0$	$m_2 = 1$	$m_2 = 2$
$m_1 = 0$	1	$1 + z$	$1 + z + \frac{1}{2} z^2$
$m_1 = 1$	$\frac{1}{1 - z}$	$\frac{1 + \frac{1}{2} z}{1 - \frac{1}{2} z}$	$\frac{1 + \frac{2}{3} z + \frac{1}{6} z^2}{1 - \frac{1}{3} z}$
$m_1 = 2$	$\frac{1}{1 - z + \frac{1}{2} z^2}$	$\frac{1 + \frac{1}{3} z}{1 - \frac{2}{3} z + \frac{1}{6} z^2}$	$\frac{1 + \frac{1}{2} z + \frac{1}{12} z^2}{1 - \frac{1}{2} z + \frac{1}{12} z^2}$

Voorbeelden

Uit een vergelijking van de tabellen 3.4 en 3.5 volgt dat de Heun-formule en de formule (3.20) en (3.21) als stabiliteitsfunctie een Padé-benadering hebben; ze zijn dus optimaal consistent. De Rosenbrock- en Calahan-formule zijn daarentegen niet optimaal consistent (zie tabel 3.4).

Stabiliteitsgebieden van Padé-polynomen

In de constructie en analyse van stabiliteitsfuncties ligt het voor de hand om eerst de stabiliteitsgebieden van de Padé-approximaties te onderzoeken.

We beschouwen eerst de klasse van Padé-polynomen ($m_1 = 0$, $m_2 = p$)

$$R(z) = \sum_{j=0}^{p} \frac{1}{j!} z^j. \tag{3.51}$$

De stabiliteitsgebieden hiervan zijn voor p = 1,2,3 en 4 in figuur 3.3 in het z = x + iy-vlak aangegeven. Aangezien deze symmetrisch zijn ten opzichte van de x-as hebben we volstaan met de contouren in het boven-halfvlak. Uit deze figuur kunnen we direct de waarden van de stabiliteitsgrenzen zen aflezen die in tabel 3.6 zijn vermeld.

Voorts volgt uit deze figuren dat in de gevallen waar de Jacobiaan uitsluitend eigenwaarden met negatief reeël deel heeft, voor voldoende kleine h_n altijd locale stabiliteit bereikt wordt.

Tabel 3.6 Stabiliteitsgrenzen van Padé-polynomen

Stabiliteitsgrens	p = 1	p = 2	p = 3	p = 4
Reeël	2	2	2.52	2.78
Imaginair	0	0	1.72	2.82
Absoluut	0	0	1.72	2.63

Een verdere toepassing van dergelijke figuren van stabiliteitsgebieden wordt gegeven door het volgende voorbeeld.

Voorbeeld

Beschouw het beginwaardeprobleem

$$y' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10^6 & -10^6-10^3 & -10^3-1 \end{pmatrix} y, \quad 0 \leq x \leq 1,$$

$$y = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad x = 0 \tag{3.52}$$

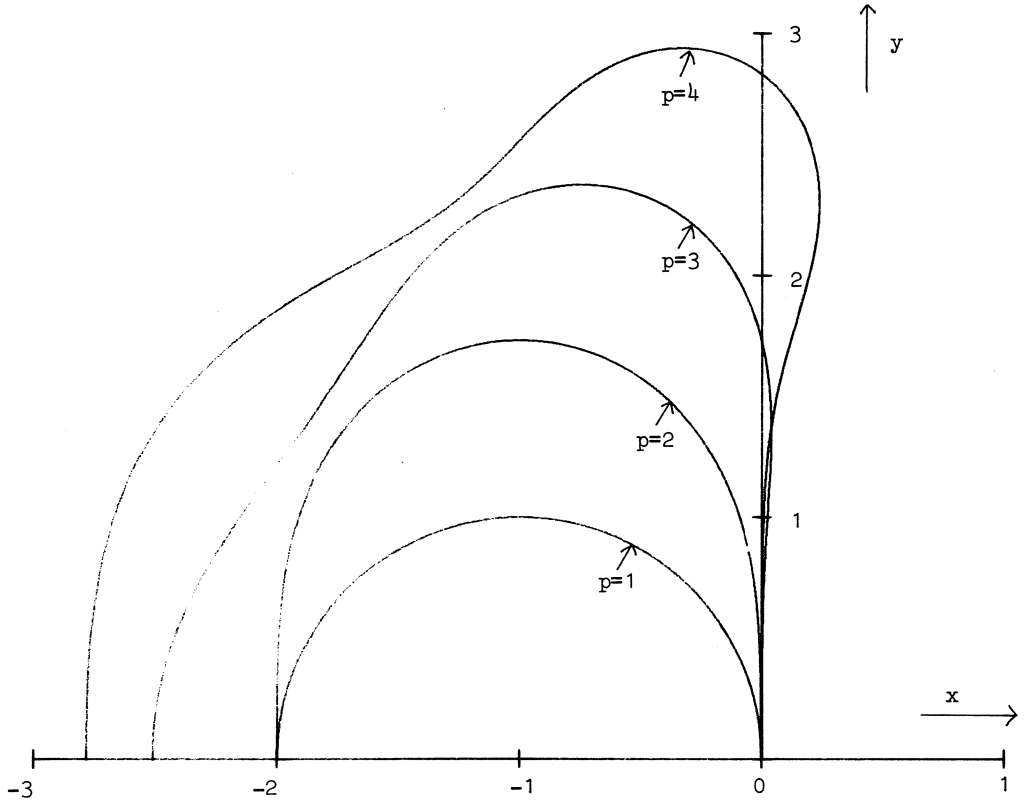


fig. 3.3 Stabiliteitsgebieden van de polynomen $R(z) = \sum_{j=0}^{p} \frac{1}{j!} z^j$ voor p = 1,2,3 en 4.

met de analytische oplossing

$$\tilde{y}(x) = e^{-x}\, y(0).$$

De eigenwaarden van de Jacobiaan van deze differentiaalvergelijking zijn:

$$\delta_1 = -1,\ \delta_2 = 1000\, e^{i\frac{2\pi}{3}},\ \delta_3 = 1000\, e^{-i\frac{2\pi}{3}}.$$

Vergelijking (3.52) is dus een typisch voorbeeld van een stijve differentiaalvergelijking. Voor stabiliteit is het nodig dat de punten $h_n\lambda_j$, $j = 1,2,3$, in het stabiliteitsgebied van de te gebruiken stabiliteitsfunctie ligt. Uit figuur 3.3 kan men afleiden dat voor de Padé-polynomen van de graad 1 tot en met 4 in benadering geldt:

$$h_n < \beta(p)10^{-3},\ \beta(1) \sim 1,\ \beta(2) \sim 2,\ \beta(3) \sim 2.5,\ \beta(4) \sim 2.64. \tag{3.53}$$

In tabel 3.7 vindt men de resultaten van enige numerieke experimenten met expliciete Taylor-formules.

Tabel 3.7 Numerieke waarden van de eerste component van de vector $\tilde{y}(1)$ verkregen met de Taylor-methode van de graad $p = 1,2,3$ en 4.

$10^3 h_n$	h_n	$p = 1$	$p = 2$	$p = 3$	$p = 4$
10	1/100	$4\ 10^{80}$	$-9\ 10^{147}$	-10^{196}	10^{234}
5	1/200	$-2\ 10^{115}$	$8\ 10^{183}$	$2\ 10^{216}$	-10^{224}
2.63	1/380	$-7\ 10^{121}$	-10^{112}	10^{18}	.36787944
2.5	1/400	$2\ 10^{119}$	$-4\ 10^{95}$	.36787944	.36787944
2	1/500	$7\ 10^{107}$	.36787969	.36787944	.36787944
1	1/1000	.36769542	.36787950	.36787944	.36787944

De analytische waarde wordt gegeven door

$$\tilde{y}(1) = e^{-1} = .3678794411\ldots$$

Vergelijken we tabel 3.7 hiermee en nemen we (3.53) in aanmerking, dan zien we dat de numerieke waarden uitermate nauwkeurig zijn zodra aan het stabiliteitscriterium voldaan is, maar nergens op lijken wanneer hieraan niet voldaan is. Dit gedrag is dus heel anders dan bij de niet-lineaire vergelijking (3.38).

Stabiliteitsgebieden van rationale Padé-benaderingen

De rationale Padé-benaderingen hebben in het algemeen een veel groter stabiliteitsgebied dan de Padé-polynomen. Men kan bijvoorbeeld bewijzen dat de "diagonale" Padé-benaderingen ($m_1=m_2$) A-stabiel zijn (zie Birkoff en Varga (1965)).

Voorbeeld

Beschouw de (2,2)-Padé-benadering

$$R(z) = \frac{1 + \frac{1}{2} z + \frac{1}{12} z^2}{1 - \frac{1}{2} z + \frac{1}{12} z^2} .$$

Het stabiliteitsgebied S bestaat alleen dan uit het gehele linker halfvlak wanneer (maximum-principe)

(3.54)
(1) $R(z)$ geen polen met negatief reëel deel heeft;
(2) $|R(iy)| \leq 1, \quad -\infty \leq y \leq \infty$;
(3) $\lim_{|z|\to\infty} |R(z)| \leq 1, \frac{\pi}{2} \leq \arg z \leq \frac{3\pi}{2}$.

Het is eenvoudig te verifiëren dat $R(z)$ aan de voorwaarden (3.54) voldoet.

Dat niet alle rationale Padé-benaderingen A-stabiel zijn toont het volgende voorbeeld aan:

Voorbeeld

Beschouw de (1,2)-Padé-benadering

$$R(z) = \frac{1 + \frac{2}{3} z + \frac{1}{6} z^2}{1 - \frac{1}{3} z} .$$

De reële stabiliteitsgrens hiervan blijkt 6 te zijn.

Polynomen met maximale reële stabiliteitsgrens

Alhoewel we beschikken over stabiliteitsfuncties die A-stabiel en consistent van elke gewenste orde p zijn, zijn de bijbehorende integratieformules in de praktijk niet altijd geschikt. Een rationale stabiliteitsfunctie brengt met zich mee dat men òf een stelsel afgebraïsche of transcendente vergelijkingen moet oplossen (impliciete methoden) òf de Jacobiaan van de differentiaalvergelijkingen moet bepalen en een vermenigvuldiging moet uitvoeren met de inverse van een matrix (semi-impliciete methoden). Behalve het feit dat de winst, verkregen door de grotere toegestane integratiestappen, weer verloren kan gaan door het extra werk, kunnen in het geval van impliciete methoden de convergentievoorwaarden van het iteratieve proces waarmee de genoemde vergelijkingen opgelost worden, soms stappen voorschrijven die nauwelijks groter zijn dan die welke genomen kunnen worden met expliciete methoden. Dit argument rechtvaardigt pogingen om polynomen te construeren met ruimere stabilitietsgebieden.

We zullen ons eerst bezighouden met het maximaliseren van de reële stabiliteitsgrens van polynomen met een orde van consistentie p, dus polynomen van de vorm

$$(3.55) \qquad R_m(z) = 1 + z + \frac{1}{2!} z^2 + \ldots + \frac{1}{p!} z^p + \beta_{p+1} z^{p+1} + \ldots + \beta_m z^m,$$

waarin we voor m_2 nu m geschreven hebben.

Voor p = 1 zijn deze polynomen bekend en werden reeds door Franklin (1959) in verband met de integratie van lineaire diffusieproblemen gebruikt. Het zijn "verschoven" Chebyshev-polynomen, gegeven door

$$(3.56) \qquad R_m(z) = T_m(1 + \frac{z}{m^2})$$

met de reële stabiliteitsgrens (zie ook figuur 3.4)

$$(3.57) \qquad \beta(m) = 2m^2.$$

Bijvoorbeeld, formule (3.13) heeft het polynoom $T_4(1 + \frac{z}{16})$ als stabiliteitsfunctie (zie tabel 3.4).

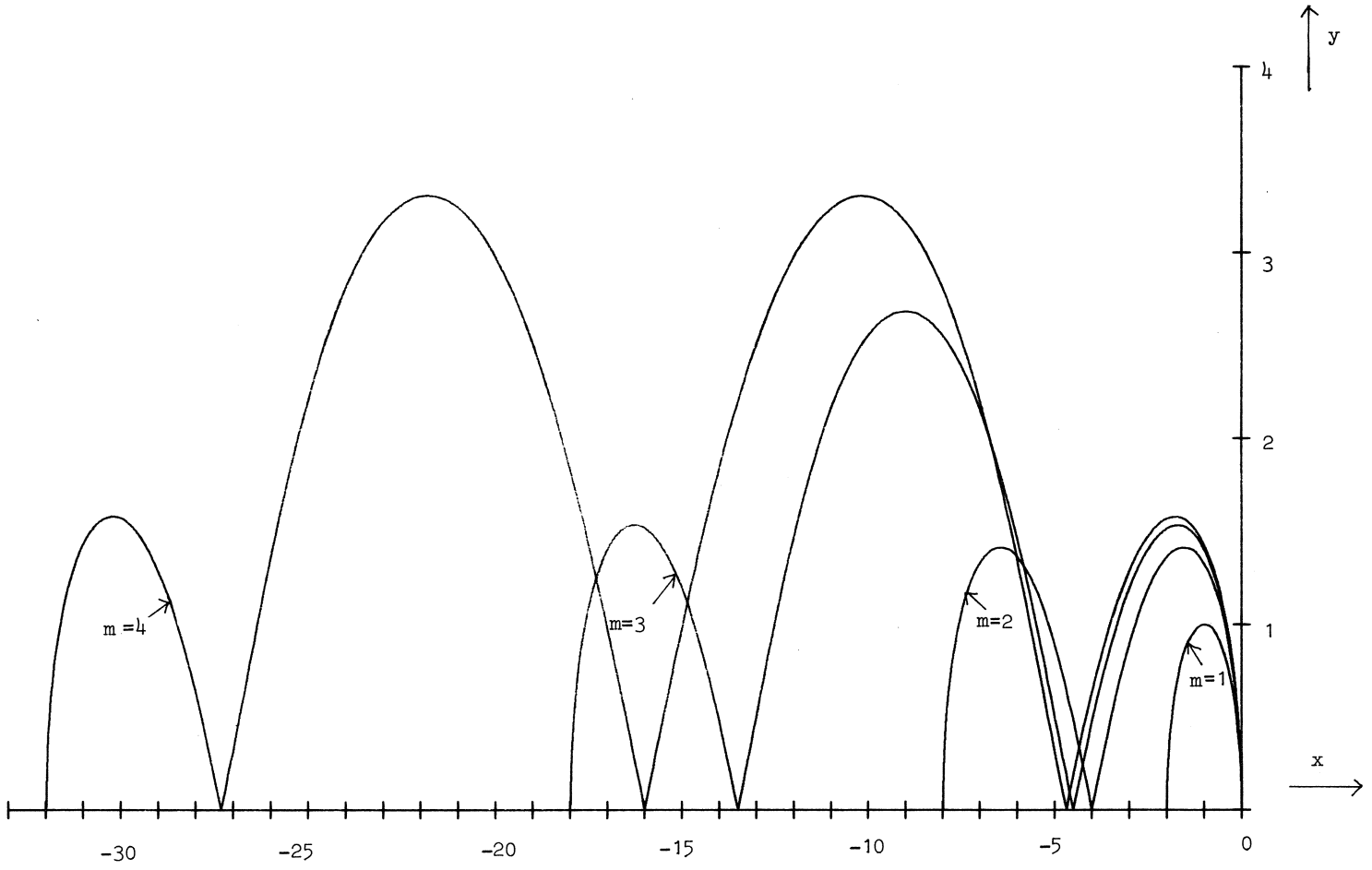


fig. 3.4. Stabiliteitsgebieden van de polynomen

Voor p = 2 is geen analytische voorstelling van het optimaliserend polynoom bekend. In een minder bekende publicatie van Lomax (1968) werd dit polynoom-probleem al onderzocht. Lomax geeft hierin een klasse van polynomen, die weliswaar een relatief grote reële stabiliteitsgrens hebben, maar niet optimaal zijn. In van der Houwen (1970a) en van der Houwen en Kok (1971) worden een groot aantal numerieke benaderingen van de maximaliserende polynomen gegeven, zowel voor p = 2 als voor p = 3 en p = 4. In dit hoofdstuk volstaan we met een tabel van de coëfficiënten van een aantal polynomen. Op de constructie van deze polynomen zal eventueel later in dit colloquium nog ingegaan worden.

Tabel 3.8. Stabiliteitspolynomen met maximale reële stabiliteitsgrens $\beta(m)$

(m,p)	$\beta(m)$	$10\beta_3$	$10^2\beta_4$	$10^3\beta_4$	$10^5\beta_6$
(3,2)	6.26	.62500000			
(4,2)	12.05	.78084485	.36084541		
(5,2)	19.45	.84608499	.55271248	.12219644	
(6,2)	28.50	.87994019	.66169168	.22176071	.2731156
(4,3)	6.03	10/6	1.8455702		
(5,3)	10.41	10/6	2.3721832	1.1118724	
(6,3)	16.05	10/6	2.6054057	1.7697690	4.284125
(5,4)	6.06	10/6	100/24	4.0869614	
(6,4)	9.97	10/6	100/24	5.3034307	24.047305

Experimenteel bleek dat voor grote waarden van m de stabiliteitsgrens $\beta(m)$, evenals in het geval p = 1, met het quadraat van m toeneemt, dus

$$\beta(m) \sim c(p)m^2. \tag{3.58}$$

Voor de evenredigheidsconstante c(p) vonden we:

$$c(2) \cong .82, \quad c(3) \cong .49, \quad c(4) \cong .34. \tag{3.59}$$

Toepassing

Beschouw het beginwaardeprobleem

$$y' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -5.10^5 & -5.10^5 - 15.10^2 & -1501 \end{pmatrix} y, \qquad 0 \leq x \leq 1,$$

(3.60)

$$y = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \qquad x = 0$$

met de analytische oplossing

$$\tilde{y}(x) = e^{-x} y(0).$$

De eigenwaarden van Jacobiaan van deze differentiaalvergelijking zijn

$$\delta_1 = -1, \quad \delta_2 = -500, \quad \delta_3 = -1000.$$

We hebben de waarde van de eerste component van de vector $\tilde{y}(1)$ numeriek berekend met de Taylor-methoden welke gegenereerd worden door de stabiliteitspolynomen gekarakteriseerd door (zie tabel 3.8)

$$(m,p) = (6,2), \quad (6,3) \quad \text{en} \quad (6,4).$$

De volgende resultaten werden verkregen:

Tabel 3.9. Numerieke waarden van de eerste component van de vector y(1) verkregen met de Taylor-methoden (m,p) = (6,2), (6,3) en (6,4).

(m,p) = (6,2) h_n = .028	(m,p) = (6,3) h_n = .016	(m,p) = (6.4) h_n = .0099
.36790226	.36787924	.36787944

De eerste 10 cijfers van de exacte oplossing worden gegeven door

$$\tilde{y}(1) = .3678794411,$$

dus de numeriek gevonden oplossing geeft respectievelijk 3, 6 en minstens 8 correcte cijfers. Vergelijken we de rekentijd van bovenstaande integratieformules met die van de standaard Runge-Kutta-formule, dan mag men, aannemende dat de rekentijd evenredig is met het aantal evaluaties van een afgeleide, stellen dat de winstfactor ongeveer 8, 4 respectievelijk 3 bedraagt. Uiteraard zal de Runge-Kutta-formule nauwkeuriger resultaten afleveren, maar wanneer een dergelijke grote nauwkeurigheid niet nodig is, heeft men bij gebruik van deze formule niet de vrijheid om met een grotere staplengte te rekenen.

Polynomen met maximale imaginaire stabiliteitsgrens

Alhoewel in de numerieke oplossing van stijve differentiaalvergelijkingen stabiliteitspolynomen met een optimale imaginaire stabiliteitsgrens niet van toepassing lijken, zijn ze toch van belang in de constructie van stabiliteitsfuncties die een stabiliteitsgebied S hebben dat een stuk van de imaginaire as bevat. Mede door het feit dat er in de numerieke analyse van symmetrische hyperbolische differentiaalvergelijkingen ook behoefte is aan een maximalisering van het imaginaire stabiliteitsinterval, is een verder onderzoek van dit optimaliseringsprobleem gerechtvaardigd.

In van der Houwen (1968) werd in verband met de numerieke oplossing van het "Noordzee-probleem" een begin gemaakt met de studie van deze maximaliserende polynomen. Gevonden werden de polynomen van de graad m = 2, 3 en 4, te weten

$$\begin{aligned} R_2(z) &= 1 + z + z^2, & \beta(2) &= 1, \\ R_3(z) &= 1 + z + \tfrac{1}{2} z^2 + \tfrac{1}{4} z^3, & \beta(3) &= 2. \\ R_4(z) &= 1 + z + \tfrac{1}{2} z^2 + \tfrac{1}{6} z^3 + \tfrac{1}{24} z^4, & \beta(4) &= 2\sqrt{2}. \end{aligned} \tag{3.61}$$

Hierin stelt $\beta(m)$ de imaginaire stabiliteitsgrens voor (de stabiliteitsgebieden zijn in figuur 3.5 gegeven). Deze polynomen zijn respectievelijk

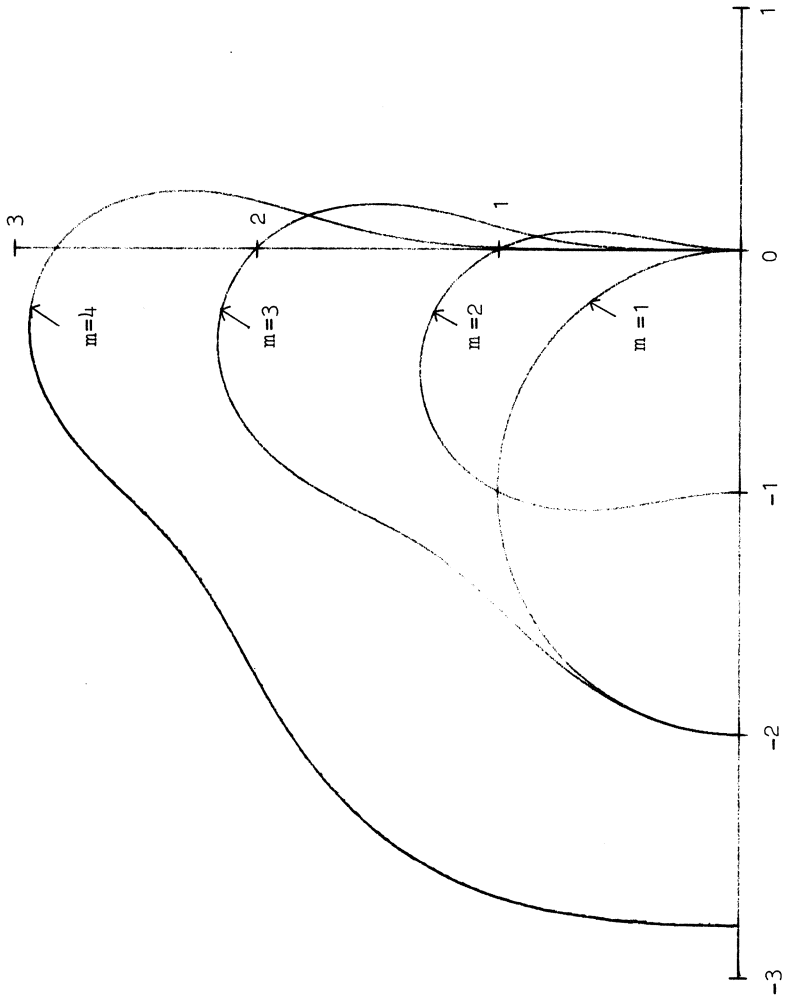


fig. 3.5. Stabiliteitsgebieden van de polynomen (3.61)

1-ste, 2-de en 4-de orde consistent. Er blijken geen optimale polynomen te bestaan met (m,p) = (3,1), (4,2) en (4,3). In van der Houwen (1969) werd een analytische uitdrukking gevonden voor het optimale polynoom van oneven graad en orde 2:

$$(3.62)\qquad T_{\frac{m-1}{2}}\left(\frac{(m-1)^2+2z^2}{(m-1)^2}\right) + 2z\,\frac{(m-1)^2+z^2}{(m-1)^2}\,U_{\frac{m-3}{2}}\left(\frac{(m-1)^2+2z^2}{(m-1)^2}\right),\quad \beta(m) = m-1.$$

Verder werd bewezen dat voor $m \geq 3$ de orde van consistentie minstens 2 is.

We zullen hier niet ingaan op de afleiding van (3.61) en (3.62). Ook dit zal in een later hoofdstuk eventueel nog behandeld worden.

Voorbeelden

De polynomen (3.61) zijn de stabiliteitsfuncties van respectievelijk de gestabiliseerde Euler-formule (3.12), de tweede gestabiliseerde Heun-formule (3.14), en de standaard Runge-Kutta-formule (3.15) (vergelijk tabel 3.4).

Polynomen met disjuncte stabiliteitsgebieden

Het beginwaardeprobleem (3.52) zou ook op stabiele wijze geïntegreerd kunnen worden met een integratieformule waarvan het stabiliteitsgebied S juist zou bestaan uit (willekeurig kleine) omgevingen van de punten $h_n\delta$, dus de punten $-h_n$, $1000h_n e^{i\frac{2\pi}{3}}$ en $1000h_n e^{-i\frac{2\pi}{3}}$. Deze overweging en het feit dat vele stijve differentiaalvergelijkingen aanleiding geven tot eigenwaardenspectra waarvan de eigenwaarden in twee of drie clusters liggen, hebben ons er toe gebracht om polynomen te construeren die een stabiliteitsgebied hebben bestaande uit omgevingen van de oorsprong en van twee vooraf gegeven toegevoegd complexe punten (zeg $z_1, \bar{z}_1$), in het linker halfvlak (zie figuur 3.6). In het bijzonder kan $z_1 = \bar{z}_1$ zijn.

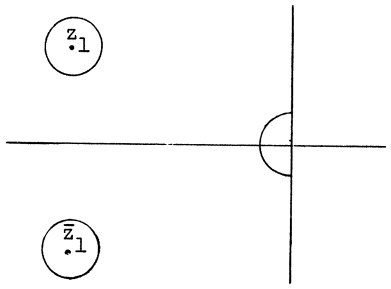


fig. 3.6. Stabiliteitsgebieden voor cluster-spectra

We zullen spreken van een *drie-clustermethode* als $z_1 \neq z_1$ en van een *twee-clustermethode* als $z_1 = \bar{z}_1$.
Het blijkt dat deze methoden voor niet al te grote cluster-diameters veel ruimere stabiliteitsvoorwaarden voorschrijven dan de methoden gebaseerd op de tot dusver besproken polynomen.

We geven hier de voornaamste resultaten van ons onderzoek van clustermethoden. Voor details zij verwezen naar van der Houwen (1971b).
We beschouwen eerst het geval dat $|z_1|$ zeer groot is. Laat $R_r(z)$ een polynoom zijn die consistent van de gewenste orde p is en het gewenste stabiliteitsgebied in de omgeving van de oorsprong heeft. We zoeken nu een polynoom $R_m(z)$ van de vorm

$$(3.63) \qquad R_m(z) = R_r(z) + z^{r+1}L_l(z), \qquad r+l+1 = m,$$

waarin L(z) een polynoom van de graad l in z is, welke zodanig is dat het stabiliteitsgebied van $R_m(z)$ in de omgeving van z_1 en $\bar{z}_1$ zo groot mogelijk is. Eenvoudigheidshalve veronderstellen we dat l oneven is. Men kan nu bewijzen (zie van der Houwen (1971c)) dat het optimale polynoom voor $|z_1| \to \infty$ bepaald wordt door de relaties

$$(3.64) \qquad R_m^{(j)}(z_1) = R_m^{(j)}(\bar{z}_1) = 0, \qquad j=0,1,2,\ldots,j_1,$$

waarin $j_1 = \frac{l-1}{2}$ als $z_1 \neq \bar{z}_1$ en $j_1 = l$ als $z_1 = \bar{z}_1$.

Voorbeeld

Voor l = 1 vinden we het polynoom

$$(3.65) \quad R_m(z) = R_r(z) + \frac{\mathrm{Im}[\bar{z}_1^m R_r(z_1)]}{|z_1|^{2(r+1)}\mathrm{Im}z_1} z^{r+1} + \frac{\mathrm{Im}[z_1^{m-1} R_r(z_1)]}{|z_1|^{2(r+1)}\mathrm{Im}z_1} z^{r+2}.$$

Stelt men $z_1 = h_n\delta_1$, waarin δ_1 het centrum van een linker cluster van eigenwaarden is, dan ontstaat een polynoom dat een drie-clustermethode genereert.
Door de limiet overgang $\bar{z}_1 \to z_1$ vinden we het polynoom dat in het twee-cluster geval toegepast kan worden.

We hebben (3.65) verder uitgewerkt voor het geval

$$r = 1, \quad R_r(z) = 1 + z.$$

Hiervoor vonden we (van der Houwen 1970b):

$$(3.66) \quad R_3(z) = 1 + z + \left[\frac{1-2\mathrm{Re}z_1}{|z_1|^2} - 4\frac{\mathrm{Re}^2 z_1}{|z_1|^2}\right] z^2 + \left[\frac{1}{|z_1|^2} + \frac{2\mathrm{Re}z_1}{|z_1|^4}\right] z^3,$$

$$z_1 = h_n\delta_1.$$

Deze stabiliteitsfunctie is alleen bruikbaar wanneer de cluster in de oorsprong geen eigenwaarden bevat die zuiver imaginair zijn, aangezien het stabiliteitsgebied van (3.66) bij de oorsprong vergelijkbaar is met dat van de methode van Euler (zie figuur 3.3). Bevat de rechter cluster wel zuiver imaginaire eigenwaarden dan zou men voor $R_r(z)$ het polynoom $1+z+z^2$ kunnen kiezen (zie figuur 3.5).

De "linker" stabiliteitsconditie van de door (3.63) en (3.64) gedefinieerde polynomen is in benadering:

$$h_n < \frac{\beta_{links}}{|\delta_l|}, \tag{3.67}$$

waarin

$$\beta_{links} = \begin{cases} -\frac{1}{r}\left[\frac{|\delta_l|^2}{2\rho_l|Im\delta_l|}\right]^{\frac{l+1}{2r}} & \text{als } \delta_l \neq \bar{\delta}_l, \\ -\frac{1}{r}\left[\frac{|\delta_l|}{\rho_l}\right]^{\frac{l+1}{r}} & \text{als } \delta_l = \bar{\delta}_l, \end{cases}$$

β_r

δ_l = het centrum van een linker eigenwaarden-cluster,

ρ_l = de straal van de linker clusters.

De "rechter" stabiliteitsconditie volgt uit het stabiliteitsgebied van $R_r(z)$.

Voorbeeld

Voor het polynoom (3.66) wordt de linker stabiliteitsvoorwaarde

$$h_n < \frac{|\delta_l|}{2\rho_l|Im\delta_l|}, \qquad Im\delta_l \neq 0$$

$$h_n < \frac{|\delta_1|}{\rho_1^2}, \qquad \mathrm{Im}\delta_1 = 0.$$

De rechter voorwaarde wordt (vergelijk het stabiliteitsgebied van de Euler-methode)

$$h_n < \frac{1}{\mathrm{Max}(|\delta_r|,\rho_1)},$$

waarin (δ_r,ρ_r) de rechter eigenwaarden-cluster voorstelt met δ_r negatief reëel.
We hebben de Taylor-methode, gegenereerd door (3.66), toegepast op het beginwaardeprobleem (3.52). De methode is voor dit probleem stabiel als $h_n < 1$. In tabel 3.10 vindt men de verkregen resultaten.

Tabel 3.10. Numerieke waarden van de eerste component van de vector $\tilde{y}(1)$ verkregen met behulp van polynoom (3.66).

h_n	$\lvert z_1\rvert$	$y(1)$	$\tilde{y}(1)$
1/2	500	.251	
1/5	200	.328	.3678
1/10	100	.349	
1/100	10	.366	

Zoals reeds is opgemerkt gelden bovenstaande resultaten voor $|z_1| \to \infty$. Wanneer $|z_1|$ kleiner wordt zal het linker stabiliteitsgebied van de polynomen (3.63), (3.64) niet meer optimaal zijn, maar men kan bewijzen dat de stabiliteitsvoorwaarden in elk geval ruimer zijn dan voorwaarde (3.67). Echter wanneer $|z_1|$ afneemt betekent dit dat h_n kleiner gekozen wordt ($z_1 = h_n\delta_1$), zodat het blijkbaar alleen van belang is om voor $|z_1| \to \infty$ de optimale polynomen te beschikken.
Wanneer men tenslotte integratiestappen kiest waarbij $|z_1|$ kleiner dan 1 wordt, dan is stabiliteit geen reden meer om bijzondere stabiliteitsfunc-

ties te gebruiken. Uit figuur 3.3 blijkt dat de Padé-polynomen ruimschoots voldoende zijn, terwijl bovendien de orde van consistentie daarbij maximaal is. Een elegante methode om de polynomen, bepaald door (3.64) waarin $|z_1| >> 1$, continu te laten overgaan in Padé-polynomen, wordt verkregen door de techniek van de exponentiële aanpassing toe te passen.

Exponentiële aanpassing

Deze techniek is in hoofdstuk 2 al geïllustreerd aan een aantal voorbeelden. We brengen in herinnering de methoden van Pope, Lawson en Liniger-Willoughby. In feite komt deze techniek er op neer dat de stabiliteitsfunctie van de integratieformule in een aantal punten van het $z = h_n\delta$-vlak aangepast wordt aan de *analytische stabiliteitsfunctie* exp z.

Stel dat men een cluster-spectrum heeft met een cluster gecentreerd in $\delta = \delta_0$. Men kan dan een stabiliteitsfunctie R(z) exponentieel hierbij aanpassen door de relaties

$$\frac{d^j}{dz^j} R(z)\Big|_{z=z_0} = e^{z_0}, \qquad j = 0,1,\ldots, \tag{3.68}$$

waarin $z_0 = h_n\delta_0$. Passen we (3.68) met $z_0 = z_1$ toe op het polynoom (3.63), waarin $j = 0,1,\ldots,\frac{l-1}{2}$ als $z_1 \neq \bar{z}_1$ en $j = 0,1,\ldots,l$ als $z_1 = \bar{z}_1$, dan ontstaan voor $|z_1| \to \infty$ dezelfde polynomen als gedefinieerd door (3.64), immers $e^{z_1} \to 0$ als $|z_1| \to \infty$; voor $|z_1| \to 0$ gaat het polynoom (3.63), (3.68) over in een m-de graads Padé-polynoom, op voorwaarde dat $R_r(z)$ een r-de graads Padé-polynoom is. Is $R_r(z)$ geen Padé-polynoom dan zou exponentieel aanpassen tot singulariteiten in de coëfficiënten van $L_l(z)$ leiden; men kan dit eenvoudig verhelpen door de coefficienten β_j van $R_r(z)$ in het gebied $|z_1| < 1$ continu naar 1/j! te laten gaan voor $z_1 \to 0$.

We zien dus dat exponentiële aanpassing op flexibele wijze de voor $|z_1| \to \infty$ optimale polynomen "plakt" aan de voor $z_1 \to 0$ optimale polynomen.

Voorbeeld

Voor l = 1 vinden we het polynoom

$$(3.65') \quad R_m(z) = R_r(z) + \frac{\mathrm{Im}\left[z_1 g(\bar{z}_1)\right]}{\mathrm{Im}\ z_1} z^{r+1} + \frac{\mathrm{Im}\ g(z_1)}{\mathrm{Im}\ z_1} z^{r+2},$$

waarin

$$g(z) = \frac{e^z - R_r(z)}{z^{r+1}}, \qquad z_1 = h_n \delta_1.$$

In het bijzonder ontstaat voor $R_r(z) \equiv 1 + z$

$$(3.66') \quad R_3(z) = 1 + z + \left[\frac{1-2\mathrm{Re}z_1}{|z_1|^2} - 4\frac{\mathrm{Re}^2 z_1}{|z_1|^4} - e^{\mathrm{Re}z_1}\frac{\sin(\mathrm{Im}z_1 - 3\phi)}{|z_1|\ \mathrm{Im}\ z_1}\right] z^2$$

$$+ \left[\frac{1}{|z_1|^2} + 2\frac{\mathrm{Re}z_1}{|z_1|^4} + e^{\mathrm{Re}z_1}\frac{\sin(\mathrm{Im}z_1 - 2\phi)}{|z_1|^2\ \mathrm{Im}\ z_1}\right] z^3,$$

waarin $\phi = \arg z_1$ (vergelijk (3.66)).

Het principe van exponentiële aanpassing is door Liniger en Willoughby toegepast op rationale stabiliteitsfuncties. Zij gebruikten de volgende stabiliteitsfuncties (zie Liniger-Willoughby (1970))

$$(3.69) \qquad R(z) = \frac{1 + \beta_1(z)}{1 - (1-\beta_1)z},$$

$$\beta_1 = -\frac{1}{z_0} - \frac{1}{e^{-z_0} - 1};$$

$$(3.70)\qquad R(z) = \frac{1 + \frac{1}{2}(1-\alpha_1)z + \frac{1}{4}(\alpha_2-\alpha_1)z^2}{1 - \frac{1}{2}(1+\alpha_1)z + \frac{1}{4}(\alpha_2+\alpha_1)z^2} ,$$

$$\alpha_1 = 2\,\frac{g(z_1) - g(z_0)}{z_1 g(z_0) - z_0 g(z_1)} ,$$

$$\alpha_2 = 2\,\frac{z_1 - z_0}{z_1 g(z_0) - z_0 g(z_1)} ,$$

$$g(z) = z^2\,\frac{1 - e^z}{e^z(2-z) - (2+z)} ;$$

$$(3.71)\qquad R(z) = \frac{1 + \frac{1}{2}(1-\alpha_1)z + \frac{1}{12}(1-3\alpha_1)z^2}{1 - \frac{1}{2}(1+\alpha_1)z + \frac{1}{12}(1+3\alpha_1)z^2} ,$$

$$\alpha_1 = \frac{1}{3z_0}\,\frac{z_0^2 + 6z_0 + 12 - e^{z_0}(z_0^2-6z_0+12)}{e^{z_0}(z_0-2) + (z_0+2)} .$$

Hierin zijn z_0 en z_1 de punten waarin de exponentiële aanpassing plaatsvindt, dus $R(z_0) = \exp(z_0)$ en, in de tweede functie, $R(z_1) = \exp(z_1)$. In de praktijk zal men, wil men complex rekenen vermijden, deze stabiliteitsfuncties gebruiken voor reële z_0 (en z_1) of complex toegevoegde punten z_0 en z_1, bijvoorbeeld $z_0 = z_1 = z_1$ als $z_1 = \bar{z}_1$ of $z_0 = z_1$ en $z_1 = \bar{z}_1$ als $z_1 \neq \bar{z}_1$. Bovenstaande functies zijn respectievelijk eerste, tweede en derde orde consistent (toepassing van tabel 3.1). Voorts bewezen Liniger en Willoughby dat deze stabiliteitsfuncties voor reële z_0 (en z_1) A-stabiel zijn; de functie (3.70) is voor complex toegevoegde z_0 en z_1 A-stabiel, mits z_0 en z_1 niet te dicht bij de imaginaire as liggen.
Tenslotte merken we op dat (3.71) uit (3.70) ontstaat door $\alpha_2 = \frac{1}{3}$ te kiezen.

Constructie van Taylor-methoden met voorgeschreven stabiliteitsfunctie en orde van consistentie

Zoals bij de introductie van de stabiliteitsfunctie in dit hoofdstuk al opgemerkt is, wordt een Taylor-methode volledig bepaald door zijn stabiliteitsfunctie, omdat de Taylor-coëfficiënten te identificeren zijn met de coëfficiënten van de stabiliteitsfunctie (zie tabel 3.4). De orde van consistentie van de stabiliteitsfunctie is dus tevens de orde van consistentie van de gegenereerde Taylor-methode.

Constructie van Runge-Kutta-methoden met voorgeschreven stabiliteitspolynoom en orde van consistentie p = 1,2,3

In tegenstelling tot de Taylor-methoden kan in het geval van de Runge-Kutta-methoden bij een gegeven stabiliteitspolynoom nog een klasse van Runge-Kutta-formules gekozen worden. Bovendien is de orde van consistentie van zo'n formule niet altijd gelijk aan de orde van consistentie van het stabiliteitspolynoom (tenzij $m \leq 2$ is of de te integreren differentiaalvergelijking lineair is).

Men kan het niet-éénduidig vastleggen van de formule door het stabiliteitspolynoom benutten door formules te kiezen welke de een of andere prettige eigenschap hebben. Wij kozen formules die een minimaal geheugengebruik vragen. We hebben getracht om in de klasse welke door het gegeven stabiliteitspolynoom gegenereerd wordt, de formules te vinden welke voorgesteld kunnen worden door het schema

$$(3.72)\begin{pmatrix} \Lambda \\ \Theta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \lambda_{10} & 0 & 0 & & 0 & 0 \\ \theta_0 & \lambda_{21} & 0 & & 0 & 0 \\ \theta_0 & 0 & \lambda_{32} & & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ \theta_0 & 0 & \cdots & \lambda_{m-1,m-2} & 0 & \\ \theta_0 & 0 & \cdots & 0 & \theta_{m-1} & \end{pmatrix}$$

Formules van dit type vragen slechts twee à drie arrays (afhankelijk van de onderlinge koppeling van de differentiaalvergelijkingen) bij een berekening op de rekenmachine (zie van der Houwen (1971b).

De karakteristieke parameters β_j en β_{jl}, waarvan de eerste 4 voor de algemene Runge-Kutta-formule gedefinieerd worden door (3.24), zijn voor het bijzondere geval (3.72) te schrijven als:

$$
\begin{aligned}
\beta_1 &= \theta_0 + \theta_{m-1}, \\
\beta_2 &= \theta_{m-1}(\theta_0+\lambda_{m-1,m-2}), \\
\beta_3 &= \theta_{m-1}\lambda_{m-1,m-2}(\theta_0+\lambda_{m-2,m-3}), \\
\beta_{31} &= \theta_{m-1}(\theta_0+\lambda_{m-1,m-2})^2.
\end{aligned}
\tag{3.24'}
$$

Zij nu p de gewenste orde van consistentie ($p \leq 3$), dan kan men met behulp van tabel 3.2 enkele Runge-Kutta-parameters expliciet berekenen. We vinden

Tabel 3.2'. Consistentievoorwaarden voor formule (3.72)

p	θ_0	θ_{m-1}	$\lambda_{m-1,m-2}$	$\lambda_{m-2,m-3}$
1	$1-\theta_{m-1}$	$1-\theta_0$		
2	$1-\theta_{m-1}$	$1-\theta_0$	$1/2\theta_{m-1}-\theta_0$	
3	1/4	3/4	5/12	17/60

We merken op dat voor p = 1 of p = 2 de parameter θ_0 nog vrij gekozen kan worden. De keuze $\theta_0 = 0$ komt in aanmerking omdat dit de integratieformule vereenvoudigt, de keuze $\theta_0 = 1/4$ komt in aanmerking omdat daarmee in het algemeen de afbreekfout kleiner wordt.

Zoals we de consistentievoorwaarden als expliciete formules voor een aantal van de Runge-Kutta-parameters geschreven hebben, zo willen we de overige parameters uitdrukken in de coëfficiënten van het voorgeschreven stabiliteitspolynoom. Daartoe bepalen we eerst het stabiliteitspolynoom van (3.72). Hiervoor vinden we

$$
R_m(z) = 1 + \beta_1 z + \beta_2 z^2 + \ldots + \beta_m z^m, \tag{3.73}
$$

waarin de coëfficiënten $\beta_1, \beta_2, \beta_3$ door middel van (3.24') in de Runge-Kutta-parameters uitgedrukt kunnen worden en de volgende coëfficiënten β_j volgens

$$\beta_j = \theta_{m-1} \prod_{l=m-j+2}^{m-1} \lambda_{l,l-1}(\theta_0+\lambda_{m-j+1,m-j}), \qquad j = 4,5,\ldots,m-1,$$

(3.74)

$$\beta_m = \theta_{m-1} \prod_{l=1}^{m-1} \lambda_{l,l-1}.$$

Identificeren we de coëfficiënten β_j met de coëfficiënten van het voorgeschreven stabiliteitspolynoom, dan kunnen uit de resulterende relaties de Runge-Kutta-parameters als functie van de nu gegeven coëfficiënten β_j gevonden worden. Tabel 3.2' kan nu uitgebreid worden tot:

Tabel 3.11. Runge-Kutta-parameters uitgedrukt in de coëfficiënten β_j van het stabiliteitspolynoom

p	θ_0	θ_{m-1}	$\lambda_{m-1,m-2}$	$\lambda_{m-2,m-3}$	$\lambda_{j,j-1}$ $j=2,\ldots,m-3$	λ_{10}
1	$1-\theta_{m-1}$	$1-\theta_0$	$\dfrac{\beta_2}{\theta_{m-1}} - \theta_0$	$\dfrac{\beta_3}{\theta_{m-1}\lambda_{m-1,m-2}} - \theta_0$	$\dfrac{\beta_{m-j+1}}{\beta_{m-j-1}} \cdot$	$\dfrac{\beta_m}{\beta_{m-1}} \cdot$
2			$\dfrac{1}{2\theta_{m-1}} - \theta_0$		$\cdot\left(1+\dfrac{\theta_0}{\lambda_{j+1,j}}\right)-\theta_0$	$\cdot\left(1+\dfrac{\theta_0}{\lambda_{21}}\right)$
3	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{5}{12}$	$\frac{17}{60}$		

Samenvattend kunnen we zeggen dat er inderdaad Runge-Kutta-formules van het type (3.72) bestaan met een orde van consistentie 1, 2 of 3 welke elk stabiliteitspolynoom met dezelfde orde van consistentie toelaten. Voor vierde orde consistente formules zij verwezen naar van der Houwen (1971b).

Voorbeelden

Formule (3.13) is een eerste orde consistentie formule van het type (3.72) met het stabiliteitspolynoom $T_4(1 + \frac{z}{16})$ (zie tabel 3.4).
De formules (3.14) zijn tweede orde consistent en hebben respectievelijk als stabiliteitspolynoom het polynoom (m,p) = (3,2) uit tabel 3.8 en het polynoom $R_3(z)$ uit (3.61).

Constructie van semi-Runge-Kuttamethoden met voorgeschreven stabiliteitsfunctie en orde van consistentie p = 1,2,3

In het algemeen zal men een semi-Runge-Kutta-methode gebruiken wanneer de evaluatie van het rechterlid van de te integreren differentiaalvergelijking erg kostbaar is in rekentijd. Met een semi-Runge-Kuttamethode kan men namelijk een redelijke orde van consistentie bereiken, terwijl relatief weinig functie-evaluaties nodig zijn.

Een-puntsformules

Voor m = 1 wordt formule (3.17) gereduceerd tot

$$(3.17') \qquad y_{n+1} = y_n + \theta_0 h_n R^{(0)}(h_n J_n) f(y_n), \quad R^{(0)}(z) = \frac{\sum_0^{m_2} \lambda_j z^j}{\sum_0^{m_1} \mu_j z^j}.$$

Het is eenvoudig te verifiëren dat voor $h_n \to 0$ geldt

$$y_{n+1} = y_n + \theta_0 h_n R^{(0)}(0) f(y_n) + \theta_0 h_n^2 R^{(0)'}(0) J_n f(y_n) +$$
$$+ \frac{1}{2} \theta_0 h_n^3 R^{(0)''}(0) J_n^2 f(y_n) + O(h_n^4),$$

waarin het accent differentiatie naar het argument van $R^{(0)}(z)$ betekent. We zien hieruit dat de eenpuntsformule (3.17') eerste orde consistent is als

$$(3.75) \qquad \theta_0 R^{(0)}(0) = 1 \quad \text{of} \quad \theta_0 \frac{\lambda_0}{\mu_0} = 1$$

en tweede orde consistent als bovendien

$$\theta_0 R^{(0)'}(0) = \frac{1}{2} \quad \text{of} \quad \theta_0\left[\frac{\lambda_1}{\mu_0} - \frac{\lambda_0\mu_1}{\mu_0^2}\right] = \frac{1}{2}\,. \tag{3.76}$$

Verder zien we dat (3.17') in het algemeen niet derde orde consistent kan zijn. Alleen wanneer f(y) lineair is en

$$\theta_0 R^{(0)''}(0) = \frac{1}{3}\,, \tag{3.77}$$

dan garanderen (3.75) - (3.77) derde orde consistentie.

De stabiliteitsfunctie van (3.17') wordt gegeven door

$$R(z) = 1 + \theta_0 z R^{(0)}(z). \tag{3.78}$$

Men kan R(z) nu gaan identificeren met een gegeven stabiliteitsfunctie en de coëfficiënten θ_0, λ_j en μ_j uitdrukken in de coëfficiënten van deze functie. In tabel 3.12 zijn een aantal resultaten opgenomen.

Tabel 3.12. Runge-Kuttaparameters uitgedrukt in de coëfficiënten van het stabiliteitspolynoom

	Polynoom	Functies van Liniger-Willoughby		
	$R_m(z) = 1 + \sum_1^m \beta_j z^j$	(3.69)	(3.70)	(3.71)
m_1	0	1	2	2
m_2	m-1	0	1	1
θ_0	1	1	1	1
λ_0	β_1	1	1	1
λ_1	β_2	0	$-\frac{1}{2}\alpha_1$	$-\frac{1}{2}\alpha_1$
λ_j	β_{j+1}, j=2,...,m-1	0	0	0
μ_0	1	1	1	1
μ_1	0	$\beta_1 - 1$	$-\frac{1}{2}(1+\alpha_1)$	$-\frac{1}{2}(1+\alpha_1)$
μ_2	0	0	$\frac{1}{4}(\alpha_2+\alpha_1)$	$\frac{1}{4}(\frac{1}{3}+\alpha_1)$
μ_j	0, $j \geq 3$	0	0	0
p	1 als $\beta_1=1$ 2 als $\beta_1=2\beta_2=1$	1	2	2

Twee-puntsformules

Voor m = 2 ontstaat uit (3.17)

$$(3.17'') \quad y_{n+1} = y_n + \theta_0 h_n R^{(0)}(h_n J) f(y_n) + \theta_1 h_n R^{(1)}(h_n J) f(y_n + \alpha h_n R^{(0)}(h_n J) f(y_n)).$$

Voor $h_n \to 0$ vinden we na enig elementair maar langdradig rekenwerk:

$$\begin{aligned} y_{n+1} = y_n &+ [\theta_0 R^{(0)}(0) + \theta_1 R^{(1)}(0)] h_n f(y_n) + \\ &+ [\theta_0 R^{(0)'}(0) + \theta_1 R^{(1)'}(0) + \alpha\theta_1 R^{(0)}(0) R^{(1)}(0)] h^2 J_n f(y_n) + \\ &+ \tfrac{1}{2}[\theta_0 R^{(0)''}(0) + \theta_1 R^{(1)''}(0) + 2\alpha\theta_1 (R^{(0)}(0) R^{(0)'}(0) R^{(1)}(0)) + \\ &\qquad - \alpha^2 \theta_1 (R^{(0)}(0) R^{(1)}(0))^2] h_n^3 J_n^2 f(y_n) + \\ &+ \tfrac{1}{2}\alpha^2 \theta_1 (R^{(0)}(0) R^{(1)}(0))^2 h_n^3 y_n''' + O(h_n^4). \end{aligned}$$

Hieruit volgt dat (3.17") consistent is van de orde 1 wanneer

$$(3.79) \qquad \theta_0 R^{(0)}(0) + \theta_1 R^{(1)}(0) = 1,$$

consistent van de orde 2 wanneer bovendien

$$(3.80) \qquad \theta_0 R^{(0)'}(0) + \theta_1 R^{(1)'}(0) + \alpha\theta_1 R^{(1)}(0) R^{(1)}(0) = \frac{1}{2},$$

en consistent is van de orde 3 wanneer behalve (3.79) en (3.80) ook

$$(3.81) \quad \begin{aligned} &\alpha^2 \theta_1 (R^{(0)}(0) R^{(1)}(0))^2 = \frac{1}{3}, \\ &\theta_0 R^{(0)''}(0) + \theta_1 R^{(1)''}(0) + 2\alpha\theta_1 (R^{(0)}(0) R^{(1)'}(0) + R^{(0)'}(0) R^{(1)}(0)) = \frac{1}{3}. \end{aligned}$$

We zullen nu een tweetal integratieformules van het type (3.17") afleiden, die consistent van de derde orde zijn.

Een expliciete formule

Kies voor de functies $R^{(0)}(z)$ en $R^{(1)}(z)$ de polynomen

$$R^{(0)}(z) = 1 + \lambda_1 z + \lambda_2 z^2$$
$$R^{(1)}(z) = 1. \tag{3.82}$$

Substitutie in (3.79) - (3.81) leidt tot de consistentie-relaties

$$\theta_0 + \theta_1 = 1,$$
$$\theta_0\lambda_1 + \alpha\theta_1 = \frac{1}{2},$$
$$\alpha^2\theta_1 = \frac{1}{3},$$
$$2\theta_0\lambda_2 + 2\alpha\theta_1\lambda_1 = \frac{1}{3}$$

en het stabiliteitspolynoom

$$R_4(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \theta_1\alpha\lambda_2 z^4.$$

Oplossing van de consistentie-relaties levert de integratie-formule (3.22). Identificeert men het stabiliteitspolynoom met een willekeurig derde orde consistent polynoom van de vierde graad, dan geeft dit voor α de voorwaarde

$$\alpha(3\alpha^2-3\alpha+1) - 6\beta_4(3\alpha^2-1)^2 = 0 \tag{3.83}$$

waarin β_4 de coefficient van z^4 in het te identificeren polynoom is. Het is eenvoudig in te zien dat deze vergelijking voor elke waarde van β_4 een reële oplossing voor α toelaat.

Een semi-impliciete formule

Kies

$$R^{(0)}(z) = \frac{1 + \nu z}{1 - \lambda z - \mu z^2},$$

$$R^{(1)}(z) \equiv 1. \tag{3.84}$$

De consistentie-voorwaarden worden

$$\theta_0 + \theta_1 = 1,$$

$$\theta_0(\nu+\lambda) + \alpha\theta_1 = \frac{1}{2},$$

$$\alpha^2\theta_1 = \frac{1}{3}$$

$$\theta_0(\lambda^2+\nu\lambda+\mu) + \alpha\theta_1(\nu+\lambda) = \frac{1}{3}$$

en de stabiliteitsfunctie wordt

$$R(z) = \frac{1 + (\theta_0+\theta_1-\lambda)z + (\theta_0\nu+\theta_1(\alpha-\lambda)-\mu)z^2 + \theta_1(\alpha\nu-\mu)z^3}{1 - \lambda z - \mu z^2}.$$

We zullen trachten deze stabiliteitsfunctie te identificeren met de functie (3.71) van Liniger-Willoughby. Dit betekent dat behalve aan de consistentie-voorwaarden ook nog voldaan moet worden aan de relaties

$$\theta_0 + \theta_1 - \lambda = \frac{1}{2}(1-\alpha_1),$$

$$\theta_0\nu + \theta_1(\alpha-\lambda) - \mu = \frac{1}{12}(1-3\alpha_1),$$

$$\alpha\nu - \mu = 0,$$

$$\lambda = \frac{1}{2}(1+\alpha_1),$$

$$\mu = -\frac{1}{12}(1+3\alpha_1).$$

Het is eenvoudig te verifiëren dat de volgende parameters een oplossing leveren voor de consistentie- en identificatie-voorwaarden:

$$\theta_0 = \frac{3\alpha^2-1}{3\alpha^2}\,, \qquad \theta_1 = \frac{1}{3\alpha^2}\,,$$

$$\nu = -\frac{1+3\alpha_1}{12}\,, \tag{3.85}$$

$$\lambda = \frac{1+\alpha_1}{2}\,, \qquad \mu = -\frac{1+3\alpha_1}{12}\,,$$

waarin α voldoet aan de vergelijking

$$\alpha_1 = \frac{-9\alpha^2+6\alpha-1}{3(2\alpha-1)(3\alpha^2-1)}\,. \tag{3.86}$$

Uit de definitie van α_1 volgt dat α_1 waarden uit het interval $[0,\frac{1}{3}]$ aanneemt. In figuur 3.7 is een grafiek voor (3.86) geschetst.

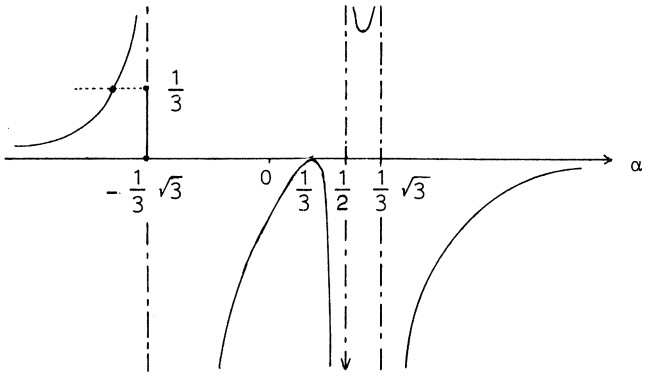


fig. 3.7. Parameter α_1 als functie van α.

Hieruit blijkt dat er voor elke $\alpha_1 \in [0,\frac{1}{3}]$ een reële waarde voor α ($\alpha < -\frac{1}{3}\sqrt{3}$) te vinden is zodanig dat aan (3.86) voldaan is. Waneer in de formule voor α_1 het punt z_0 naar nul gaat (cf. (3.71)), dan nadert α_1 tot 0 en α tot $-\infty$. In numerieke berekeningen zijn groot negatieve waarden van de parameters echter ongewenst en men doet er dan beter aan om voor bijvoorbeeld $\alpha_1 < 0.005$ ($\alpha<-10$) te stellen $\alpha_1 = 0$, waardoor $\alpha = \frac{1}{3}$ een oplossing

van (3.86) levert. De integratieformule gaat dan over in formule (3.21) met de (2,2)-Padé-benadering als stabiliteitsfunctie.

Van de hierboven afgeleide tweepuntsformules zijn nog geen numerieke resultaten beschikbaar aangezien ze van zeer recente datum zijn. In feite zijn ze ontwikkeld naar aanleiding van dit colloquium.

REFERENTIES

Birkoff, G. and R.S. Varga (1965), Discretization errors for well-set Cauchy problems I, J. Math. and Phys. 44, 1.

Butcher, J.C. (1964a), Implicit Runge-Kutta processes, Math. Comp. 18, 50.

Butcher, J.C. (1964b), Integration processes based on Radau quadrature formulas, Math. Comp. 18,233.

Calahan, D.A. (1968), A stable, accurate method of numerical integration for non-linear systems, Proc. IEEE 56, 744.

Dahlquist, G. (1963), A special stability problem for linear multistep methods, BIT 3, 27.

Franklin, J.N. (1959), Numerical stability in digital and analogue computation for diffusion problems, J. Math. and Phys. 37, 305.

Henrici, P. (1962), Discrete variable methods in ordinary differential equations, Wiley, New York.

Heun, K. (1900), Neue Methode zur approximativen Integration der Differentialgleichungen einer unabhängigen Veränderlichen, Zeitschr. Math. und Phys. 45, 435.

Houwen, P.J. van der (1968), Finite difference methods for solving partial differential equations, MC Tract 20, Math. Centrum, Amsterdam.

Houwen, P.J. van der (1969), Difference schemes with complex time steps, Report MR 105, Mathematisch Centrum, Amsterdam.

Houwen, P.J. van der (1970a), One-step methods for linear initial value problems I, Polynomial methods, Report TW 119, Mathematisch Centrum, Amsterdam.

Houwen, P.J. van der (1970b), One-step methods for linear initial value problems II, Applications to stiff equations, Report TW 122, Mathematisch Centrum, Amsterdam.

Houwen, P.J. van der (1971a), Stabilized Runge-Kutta methods with limited storage requirements, Report TW 124/71, Mathematisch Centrum, Amsterdam.

Houwen, P.J. van der, P. Beentjes, K. Dekker, E. Slagt, (1971b), One-step methods for linear initial value problems III, Numerical examples, Report TW 130, Mathematisch Centrum, Amsterdam.

Houwen, P.J. van der, (1971c), A survey of stabilized Runge-Kutta methods, MC Tract 37, Ch. 5, Mathematisch Centrum, Amsterdam.

Houwen, P.J. van der, J. Kok, (1971d), Numerical solution of a minimax problem, Report TW 123, Mathematisch Centrum, Amsterdam.

Kutta, W. (1901), Beitrag zur näherungsweisen Integration totaler Differentialgleichungen, Zeitschr. Math. und Phys., 46, 435.

Liniger, W. and R.A. Willoughby, (1970), Efficient integration methods for stiff systems of ordinary differential equations, SIAM J., Numer. Anal. 7, 47.

Lomax, H. (1968), On the construction of highly stable, explicit, numerical methods for integrating coupled ordinary differential equations with parasitic eigenvalues, NASA Technical Note, NASA TN D - 4547.

Rosenbrock, H.H. (1963), Some general implicit processes for the numerical solution of differential equations. Comput. J., 5, 329.

Runge, C. (1895), Über die numerische Auflösung von Differentialgleichungen, Math. Ann., 46, 166.

4. Lineaire meerstapsmethoden

In dit hoofdstuk zullen we methoden behandelen welke bij het berekenen van elke stap uit het integratieproces gebruik maken van informatie, die verkregen is bij het berekenen van een aantal vorige stappen. We zullen ons echter beperken tot die methoden waarbij, voor het oplossen van de differentiaalvergelijking

$$y'(x) = f(x,y), \tag{4.1}$$

gebruik gemaakt wordt van een lineaire vorm in y_i en y'_i. Een dergelijke lineaire formule schrijven we, voor een k-stapsmethode, in de algemene vorm

$$L(y) \equiv \sum_{i=0}^{k} \alpha_i\, y_{n-i} + \beta_i\, y'_{n-i} = 0. \tag{4.2}$$

De coëfficienten α_i en β_i zullen in het algemeen afhankelijk zijn van de steunpunten x_i. Wanneer echter de steunpunten equidistant gekozen worden, zijn α_i en β_i onafhankelijk van $\{x_n\}$. In dat geval spreekt men van *lineaire meerstapsmethoden met vaste staplengte*. Wanneer de steunpunten niet equidistant liggen, spreekt men van een methode *met variabele staplengte*.

Is in (4.2) de parameter β_0 gelijk aan nul, dan kan y_n direkt berekend worden uit y_i en y'_i $(n-k\leq i<n)$; (4.2) heet dan een *expliciete* of *open* formule. Is $\beta_0 \neq 0$ dan heet (4.2) een *impliciete* of *gesloten* formule.

In het algemeen wordt bij het berekenen van één stap de formule (4.2) een aantal malen gebruikt. Bijvoorbeeld wordt dan met een open formule eerst een waarde voor y_n "voorspeld", welke daarna één of meerdere malen met een gesloten formule wordt verbeterd (de zogenaamde *predictor-corrector* (P.C.) methoden).

Methoden met vaste staplengte

De begrippen consistentie, convergentie en stabiliteit zullen eerst besproken worden voor methoden met vaste staplengte. In dergelijke gevallen kunnen we met de formule (4.2) een operator L_h associëren van de vorm

$$L_h(y(x)) = \sum_{i=0}^{k} \alpha_i\, y(x-hi) + h\, \beta_i\, y'(x-hi). \tag{4.3}$$

De *orde van consistentie* van formule (4.2) wordt nu gedefinieerd als het grootste getal p waarvoor geldt

$$L_h(y(x)) = O(h^{p+1}) \tag{4.4}$$

voor iedere p + 1 maal continu differentieerbare oplossing y(x) van (4.1).

We kunnen met behulp van (4.3) en (4.4) op eenvoudige wijze voorwaarden opstellen voor de constanten α_i en β_i wanneer we de functie y(x) in een Taylorreeks ontwikkelen:

$$L_h(y(x)) = \sum_{r=0}^{p+1} C_r\, y^{(r)}(x)h^r + O(h^{p+2}) \tag{4.5}$$

met

$$\begin{aligned} C_0 &= \sum_{i=0}^{k} \alpha_i \,, \\ C_r &= \sum_{i=0}^{k} \frac{(-i)^r}{r!}\alpha_i + \sum_{i=0}^{k} \frac{(-1)^{r-1}}{(r-1)!}\beta_i \qquad (r>0). \end{aligned} \tag{4.6}$$

Voor p-de orde consistentie is derhalve noodzakelijk

$$C_r = 0 \qquad (0 \leq r \leq p).$$

De coëfficiënten α_i en β_i zijn bepaald op een constante factor na. Wanneer we normaliseren door bijvoorbeeld $\sum_{i=0}^{k} \beta_i = 1$ te kiezen, is C_{p+1} bruikbaar als foutconstante (d.i. een maat voor de nauwkeurigheid binnen de klasse van formules met dezelfde orde). Bij de constructie van een k-stapsmethode hebben we blijkbaar 2k + 1 vrijheden om α_i en β_i te kiezen. Hiermee is het mogelijk om aan (4.6) te voldoen voor $0 \leq r \leq 2k$. De aldus ontstane formule is dan van de orde 2k. Voor k > 2 voldoet deze formule echter niet aan de stabiliteitsvoorwaarden (zie stelling (4.3)).

Voorbeeld

De constructie van een 4^{de} orde 2-stapsformule komt neer op het oplossen

van (4.6) voor k = 2, $0 \leq r \leq 4$. Het stelsel vergelijkingen luidt dan

$$\begin{aligned}
\alpha_0 + \alpha_1 + \alpha_2 &= 0 \ , \\
\alpha_1 + 2\alpha_2 &= \beta_0 + \beta_1 + \beta_2 \ , \\
\alpha_1 + 4\alpha_2 &= 2(\beta_1+2\beta_2) \ , \\
\alpha_1 + 8\alpha_2 &= 3(\beta_1+4\beta_2) \ , \\
\alpha_1 + 16\alpha_2 &= 4(\beta_1+8\beta_2) \ .
\end{aligned}$$

De oplossing luidt:

$$\alpha_0 = -\alpha_2 = \alpha \qquad \alpha_1 = 0$$

$$\beta_0 = \beta_2 = \tfrac{1}{3}\alpha \qquad \beta_1 = \tfrac{4}{3}\alpha \ .$$

Het resultaat is de bekende Milne-Simpson formule

$$y_n = y_{n-2} + \tfrac{1}{3}(y'_n + 4y'_{n-1} + y'_{n-2}). \tag{4.7}$$

De polynomen $\rho(\zeta)$ en $\sigma(\zeta)$

Een aantal eigenschappen van lineaire meerstapsmethoden met variabele staplengte kan eenvoudig geformuleerd worden door gebruik te maken van de polynomen

$$\rho(\zeta) = \sum_{i=0}^{k} \alpha_i \zeta^{k-i} \quad \text{en} \quad \sigma(\zeta) = \sum_{i=0}^{k} \beta_i \zeta^{k-i} \ . \tag{4.8}$$

Stelling 4.1

De volgende beweringen zijn equivalent:

(1) L is consistent van de orde p.

(2) $$L_h(e^{\lambda x}) = C_{p+1}(h\lambda)^{p+1} e^{\lambda x} + O(h^{p+2}).$$

(3) $\qquad \rho(1+z) + \log(1+z)\ \sigma(1+z) = C_{p+1}\ z^{p+1} + O(z^{p+2}).$

Bewijs

(i) We bewijzen eerst de equivalentie van (1) en (2).
Ontwikkeling van $e^{\lambda x}$ in een Taylorreeks geeft, analoog aan (4.5),

(4.9) $\qquad L_h(e^{\lambda x}) = \sum_{r=0}^{\infty} C_r\ (h\lambda)^r\ e^{\lambda x}\ .$

Enerzijds volgt nu uit (4.4) dat (1) equivalent is met

(4.10) $\qquad C_r = 0 \quad 0 \leq r \leq p, \quad C_{p+1} \neq 0\ .$

Anderzijds volgt uit (4.9) dat (4.10) equivalent is met (2).

(ii) De equivalentie van (2) en (3) volgt uit

$$L_h(e^{\lambda x}) = \sum_{i=0}^{k} \alpha_i\ e^{\lambda x}\ e^{-i\lambda h} + \lambda h\ \beta_i\ e^{\lambda x}\ e^{-i\lambda h}$$

$$= e^{\lambda x - k\lambda h}\ [\rho(e^{\lambda h}) + \lambda h\ \sigma(e^{\lambda h})]\ ,$$

zodat

$$\rho(e^{\lambda h}) + \lambda h\ \sigma(e^{\lambda h}) = e^{-\lambda(x-kh)}\ [C_{p+1}(h\lambda)^{p+1}\ e^{\lambda x} + O(h^{p+2})]$$

$$= e^{kh\lambda}\ C_{p+1}(h\lambda)^{p+1} + e^{kh\lambda}\ O(h^{p+2})$$

$$= C_{p+1}(h\lambda)^{p+1} + O(h^{p+2}).$$

De substitutie $e^{\lambda h} = 1 + z$ levert direkt (3).
Op dezelfde wijze volgt (2) uit (3).

Gevolg

Ontwikkelen we $\rho(1+z) + \log(1+z)\ \sigma(1+z)$ in een machtreeks, dan blijkt dat voor formules van een bepaalde orde de polynomen ρ en σ aan zekere voorwaar-

den moeten voldoen.

Tabel 4.1. Relaties tussen ρ en σ voor methoden van orde p.

orde	relatie
$p \geq 0$	$\rho(1) = 0$
$p \geq 1$	$\rho'(1) + \sigma(1) = 0$
$p \geq 2$	$\rho''(1) + 2\sigma'(1) - \sigma(1) = 0$

Toepassing

De relatie $\rho(1+z) + \log(1+z)\,\sigma(1+z) = O(z^{p+1})$ kan gebruikt worden om methoden met een maximale orde te construeren als één van de polynomen ρ of σ gegeven is.

Voorbeeld

Stel dat we een 2-stapsformule willen construeren van de vorm

$$y_n = y_{n-1} + h(\beta_1 y'_{n-1} + \beta_2 y'_{n-2}).$$

Aan de voorwaarde $\sum_{i=0}^{k} \alpha_i = \rho(1) = 0$ is voldaan, zodat we β_1 en β_2 zodanig kunnen bepalen dat de methode 2-de orde consistent is.

$$\rho(\zeta) = -\zeta^2 + \zeta, \quad \sigma(\zeta) = \beta_1\zeta + \beta_2$$

$$\rho(1+z) = -(1+z)^2 + (1+z) = -z^2 - z$$

$$\sigma(1+z) = \beta_1 + \beta_1 z + \beta_2$$

$$\log(1+z) = z - \frac{z^2}{2} + O(z^3)$$

zodat β_1 en β_2 moeten voldoen aan

$$-z - z^2 + \beta_1 z + \frac{\beta_1}{2} z^2 + \beta_2 z - \frac{\beta_2}{2} z^2 = O(z^3).$$

Hieruit volgt $\beta_1 = +\frac{3}{2}$ en $\beta_2 = -\frac{1}{2}$.
De gezochte (expliciete) 2-de orde 2-stapsmethode luidt dus

$$y_n = y_{n-1} + \frac{h}{2}(3y'_{n-1} - y'_{n-2}).$$

Dit is de 2-de orde Adams-Bashforth formule.

Principale en parasitaire wortels

Als we de modelvergelijking $y' = \delta y$ substitueren in de meerstapsmethode (4.2) krijgen we

$$(4.11) \qquad \sum_{i=0}^{k} (\alpha_i + h\delta\, \beta_i)\, y_{n-i} = 0 .$$

De oplossing van deze differentievergelijking luidt

$$(4.12) \qquad y_n = \sum_{i=0}^{k} A_i\, \xi_i^n$$

waarin ξ_i $(1 \leq i \leq k)$ de wortels voorstellen van de vergelijking

$$(4.13) \qquad \sum_{i=1}^{k} (\alpha_i + h\delta\, \beta_i)\, \xi^{k-i} = \rho(\xi) + h\delta\, \sigma(\xi) = 0 .$$

Wanneer meervoudige wortels voorkomen, leveren deze wortels bijdragen tot de som (4.12), welke behalve van de vorm $A_i\xi_i^n$, ook van de vorm $A_i n\xi_i^n$, $A_i n^2\xi_i^n$, ... zijn. De constanten A_i worden bepaald door de beginwaarden y_i $(0 \leq i < k)$ van de differentievergelijking.

Uit stelling (4.1) volgt dat, voor kleine waarden van $h\delta$, er een wortel van (4.13) zal zijn, welke $e^{\delta h}$ benadert. Deze wortel (de *principale wortel*) zorgt er voor dat de oplossing van de differentievergelijking (4.11) een benadering kan zijn voor de oplossing $y_n = Ce^{nh\delta}$ van de differentiaalvergelijking $y' = \delta y$. De overige wortels van (4.13) (de *parasitaire wortels*) dienen zodanig te zijn dat de benadering van de oplossing niet ongunstig beïnvloed wordt.

Stabiliteit

De parasitaire wortels zijn in het bijzonder van belang wanneer we de stabiliteit van een methode beschouwen. Laat y_i^* $(0 \leq i \leq n)$ de waarde zijn die ontstaat wanneer y_i verstoord wordt, bijvoorbeeld door afrondingsfouten. We definiëren dan de *numerieke fout*

$$\varepsilon_i = y_i^* - y_i \ . \tag{4.14}$$

Wanneer we aannemen dat bij het toepassen van de formule $L(y) = 0$ geen fouten optreden, d.w.z. dat y_n^* exact berekend wordt uit y_i^*, dan zullen de fouten ε_i $(n-k \leq i < n)$ toch een verstoring ε_n veroorzaken in y_n.

Voor de modelvergelijking

$$y' = \delta y + f(x) \tag{4.15}$$

kunnen we, vanwege de lineariteit in y, de vergelijking opstellen waaruit ε_n volgt

$$L(y^*) - L(y) = L(\varepsilon) = 0$$

ofwel

$$\sum_{i=0}^{k} (\alpha_i + h\delta \beta_i) \varepsilon_{n-i} = 0 \tag{4.16}$$

De oplossing van deze differentievergelijking heeft weer de vorm (4.13). We kunnen nu direkt de volgende stelling formuleren

Stelling 4.2

De numerieke fout, bij het toepassen van (4.2), zal toenemen wanneer de vergelijking

$$\rho(\xi) + h\delta \sigma(\xi) = 0$$

een wortel heeft welke in absolute waarde groter dan 1 is, of wanneer een

wortel met absolute waarde gelijk aan 1 meervoudig is.

In het algemeen eist men voor stabiele problemen (Re $\delta \leq 0$) dat de parasitaire wortels in absolute waarde kleiner dan 1 zijn (*absolute stabiliteit*). Bij het oplossen van instabiele problemen (Re $\delta > 0$) is het van belang dat alle parasitaire wortels kleiner zijn dan de principale (*relatieve stabiliteit*).

Stabiliteit voor $h \to 0$ (stabiliteit in de zin van Dahlquist)

Van een methode zullen we minstens moeten eisen dat de numerieke fout niet toeneemt wanneer we de methode toepassen met kleine staplengte. Aangezien de wortels van een polynoom continue functies zijn van de coëfficiënten van dat polynoom, zullen de wortels van $\rho(\xi) + h\delta\,\sigma(\xi) = 0$ voor $h\delta \to 0$ naderen tot de wortels van $\rho(\xi) = 0$ mits de graad van $\sigma(\xi)$ kleiner dan of gelijk aan de graad van $\rho(\xi)$ is.

Definitie. Een methode heet *stabiel in de zin van Dahlquist* wanneer het bijbehorende polynoom $\rho(\xi)$ geen nulpunten heeft welke in absolute waarde groter zijn dan 1 en wanneer de nulpunten met absolute waarde gelijk aan 1 enkelvoudige multipliciteit bezitten.

We moeten opmerken dat, bij een methode met orde $p \geq 0$, de vergelijking $\rho(\xi) = 0$ minstens één wortel heeft die gelijk is aan 1, namelijk de principale wortel (zie tabel 4.1).
Een methode heet *sterk stabiel* wanneer van het polynoom $\frac{\rho(\xi)}{1-\xi}$ alle nulpunten absoluut kleiner dan 1 zijn; bezit dit polynoom ook nulpunten op de eenheidscirkel, dan heet de methode *zwak stabiel*.

Stelling 4.3 [Dahlquist (1956)]

De orde van een stabiele, lineaire k-stapsmethode is voor oneven k maximaal k + 1 en voor even k maximaal k + 2. Bij een stabiele methode van orde k + 2 liggen alle nulpunten van $\rho(\xi)$ op de eenheidscirkel.

Bewijs

Zie Henrici p. 229-232.

Convergentie

Een methode heet *convergent* wanneer voor alle oplossingen $y(x)$ van de differentiaalvergelijking welke een voldoend aantal malen differentieerbaar zijn, geldt: $y_i \to y(x_i)$ voor $h \to 0$. Alhoewel consistentie aangeeft in welke mate de formule $L(y) = 0$ de werking van de differentiaalvergelijking benadert voor $h \to 0$, impliceert consistentie nog geen convergentie.

Stelling 4.4

Noodzakelijke en voldoende voorwaarden voor convergentie van een lineaire meerstapsmethode zijn:

(i) stabiliteit in de zin van Dahlquist;

(ii) consistentie van orde $p \geq 1$.

Bewijs

Zie Henrici p. 218, 224, 242-246.

Opmerking. De voorwaarden van convergentie zijn eenvoudig uit te drukken in voorwaarden voor de polynomen ρ en σ:

(i) de wortels van $\rho(\xi) = 0$ moeten binnen of op de eenheidscirkel liggen;

(ii) $\rho(1) = 0$;

(iii) $\rho'(1) + \sigma(1) = 0$.

Stabiliteit voor eindige staplengte

Voor methoden welke bestaan uit het slechts éénmaal toepassen van de formule $L(y) = 0$ kan direkt aangegeven worden voor welke waarden van h stabiliteit gegarandeerd is. Hiertoe beschouwen we de modelvergelijking

(4.15), waarin δ een complexe scalar voorstelt. De wortels van de vergelijking in ξ

$$\rho(\xi) + h\delta\,\sigma(\xi) = 0$$

dienen nu binnen de eenheidscirkel te liggen. Alle punten $h\delta$ waarvoor dit te realiseren is, vormen een gebied in het $h\delta$-vlak, dat begrensd wordt door die punten $h\delta$ waarvoor geldt

$$h\delta = -\frac{\rho(e^{i\phi})}{\sigma(e^{i\phi})}, \qquad 0 \leq \phi \leq 2\pi.$$

Een stabiliteit gebied is met behulp van deze formule eenvoudig te construeren. De stabiliteit van P.C.-methoden, waarbij meermalen een formule $L(y) = 0$ toegepast wordt, zal in een volgende paragraaf behandeld worden.

Om diverse eigenschappen met betrekking tot stabiliteit voor eindige staplengte aan te geven zijn verschillende begrippen ingevoerd.

Definitie. Een methode heet *A-stabiel*, wanneer hij stabiel is voor iedere $h\delta$ met $\mathrm{Re}(h\delta) < 0$.

Definitie. Een methode heet *A(α)-stabiel*, wanneer hij stabiel is voor iedere $h\delta$ met $|\arg(-h\delta)| < \alpha$.

Definitie. Een methode heet *"stiffly-stable"*, wanneer hij stabiel is voor iedere $h\delta$ met $\mathrm{Re}(h\delta) < M$ en bovendien in de rechthoek $\{M \leq \mathrm{Re}(h\delta) < 0,\ -\alpha < \mathrm{Im}(h\delta) < \alpha\}$.

We noemen nu enige stellingen die relevant zijn voor het oplossen van stijve differentiaalvergelijkingen.

Stelling [Dahlquist (1963)]

Een expliciete k-stapsformule $L(y)$ kan niet A stabiel zijn.

Stelling [Dahlquist (1963)]

De maximale orde van een A-stabiele lineaire meerstapsmethode bedraagt 2. De A-stabiele 2-de orde methode met de kleinste foutconstante is de trapeziumregel.

Stelling [Widlund (1967)]

Voor alle $\alpha \in [0,\pi/2)$ bestaan er A(α)-stabiele p-de orde k-stapsmethodes met $k = p = 3$ en $k = p = 4$.

Voor het bewijs van deze stellingen zij verwezen naar de oorspronkelijke literatuur.

Meerstapsmethoden als proces

Het uitvoeren van één stap van het integratie proces kan gezien worden als het toepassen van een operator E, welke de informatie, die gebruikt wordt bij het uitvoeren van die stap, overvoert in de informatie welke beschikbaar gesteld wordt nadat de stap is uitgevoerd. Deze informatie, die beschikbaar is na de n-de stap, geven we aan met een vector $\vec{y}_n$. Het uitvoeren van een stap wordt nu weergegeven door

$$\vec{y}_{n+1} = E(\vec{y}_n).$$

We gaan het deelproces, dat bestaat uit het eenmaal toepassen van deze operator E (hetzij met behulp van een formule $L(y) = 0$, hetzij met behulp van een predictor-corrector paar), nu nader beschouwen.

Stelling 4.8

Met iedere open of gesloten formule (4.2) en met iedere predictor-corrector-methode kan een proces geassocieerd worden van de volgende vorm:

$$
\begin{aligned}
{}_0\vec{y}_n &= B\,\vec{y}_{n-1} \\
{}_{(m+1)}\vec{y}_n &= {}_m\vec{y}_n + \vec{c}(hf({}_my_n) - h\,{}_my'_n) \\
\vec{y}_n &= {}_M\vec{y}_n \ .
\end{aligned}
\tag{4.17}
$$

Hierin is

$$\vec{y}_n = (y_n, y_{n-1}, \ldots, y_{n-k}, hy'_n, hy'_{n-1}, \ldots, hy'_{n-k})^T,$$

$${}_m\vec{y}_n = ({}_my_n, y_{n-1}, \ldots, y_{n-k}, h\,{}_my'_n, hy'_{n-1}, \ldots, hy'_{n-k})^T,$$

${}_my_n$ is de m-de versie van y_n,

$$\vec{c} = (\beta^*_0, 0, \ldots, 0, 1, 0, \ldots, 0)^T,$$

$$
B = \begin{pmatrix}
\alpha_1, \ldots, \alpha_k, & \beta_1, \ldots, \beta_k \\
1 \quad 0 & \\
\quad \ddots & 0 \\
0 \quad 1 \ 0 & \\
\gamma_1, \ldots, \gamma_k, & \delta_1, \ldots, \delta_k \\
 & 1 \quad 0 \\
0 & \quad \ddots \\
 & 0 \quad 1 \ 0
\end{pmatrix} .
$$

Bewijs

De werking van het proces is als volgt:

$$
\begin{aligned}
{}_1y_n &= {}_0y_n + \beta^*_0\,(hf({}_0y_n) - h\,{}_0y'_n) \\
&= \sum_{i=1}^{k} \{(\alpha_i - \beta^*_0\,\gamma_i)y_{n-i} + (\beta_i - \beta^*_0\,\delta_i)h\,y'_{n-i}\} + \beta^*_0\,hf({}_0y_n) \\
&= \sum_{i=1}^{k} (\alpha^*_i\,y_{n-i} + \beta^*_i\,h\,y'_{n-i}) + \beta^*_0\,hf({}_0y_n) \ .
\end{aligned}
\tag{4.18}
$$

(De laatste regel definieerd in feite α^*_i en β^*_i.)

$$
\begin{aligned}
h\ {}_1y'_n &= hf({}_0y_n) \\
{}_{(m+1)}y_n &= {}_my_n + \beta_0^*(hf({}_my_n) - h\ {}_my'_n) \qquad (m>0) \\
&= {}_1y_n + \beta_0^*(hf({}_my_n) - h\ {}_1y'_n) \\
&= \sum_{i=1}^{k} (\alpha_i^*\, y_{n-i} + h\,\beta_i^*\, y'_{n-i}) + \beta_0^*\, hf({}_my_n) \\
h{}_{(m+1)}y'_n &= hf({}_my_n).
\end{aligned} \tag{4.19}
$$

We bewijzen de stelling nu in drie delen.

(i) Wanneer $\beta_0^* = 0$ worden (4.18) en (4.19)

$$
\begin{aligned}
{}_my_n &= \sum_{i=1}^{k} (\alpha_i\, y_{n-i} + \beta_i\, h\, y'_{n-i}) \\
h\ {}_my'_n &= hf({}_0y_n) \qquad\qquad m \geq 1.
\end{aligned} \tag{4.20}
$$

Na een correctieslag (m=1) verandert de waarde van y_n niet meer. Het proces komt blijkbaar overeen met het toepassen van een enkele expliciete meerstapsformule. De waarden van γ_i en δ_i zijn irrelevant.

(ii) Wanneer het proces, met $\beta_0^* \neq 0$, wordt voortgezet tot dat convergentie bereikt is, geldt

$$
\begin{aligned}
y_n &= \sum_{i=1}^{k} (\alpha_i^*\, y_{n-i} + \beta_i^*\, h\, y'_{n-i}) + \beta_0^*\, h\, y'_n \\
h\ y'_n &= hf(y_n).
\end{aligned} \tag{4.21}
$$

Het proces komt overeen met een impliciete formule. De voorwaarde voor convergentie van het proces volgt direkt uit (4.19):

$$
|\beta_0^*\, h\, f_y(y_n)| < 1. \tag{4.22}
$$

(iii) Laat een willekeurig predictor-corrector paar gegeven zijn

$$(4.23) \qquad {}_0y_n = \sum_{i=1}^{k} (\alpha_i\, y_{n-i} + \beta_i\, h\, y'_{n-i})$$

$$(4.24) \qquad {}_{(m+1)}y_n = \beta_0^*\, hf({}_my_n) + \sum_{i=1}^{k} (\alpha_i^*\, y_{n-i} + \beta_i^*\, h\, y'_{n-i}).$$

Zijn γ_i en δ_i gedefinieerd door $\alpha_i - \beta_0^*\,\gamma_i = \alpha_i^*$ en $\beta_i - \beta_0^*\,\delta_i = \beta_i^*$, dan volgt uit (4.23) en (4.24)

$$(4.25) \qquad {}_1y_n - {}_0y_n = \beta_0^*\,(hf({}_0y_n) - h\,{}_0y'_n)$$

met

$$(4.26) \qquad h\,{}_0y_n = \sum_{i=1}^{k} (\gamma_i\, y_{n-i} + \delta_i\, h\, y'_{n-i}).$$

Bovendien volgt uit (4.24) voor $m > 0$

$$(4.27) \qquad {}_{(m+1)}y_n - y_{n,(m)} = \beta_0^*\,(hf({}_my_n) - h\,{}_my'_n)$$

met

$$(4.28) \qquad h\,{}_my'_n = hf({}_{(m-1)}y_n)\ .$$

De formules (4.23), (4.25), (4.27), (4.26) en (4.28) zijn juist equivalent met het proces (4.17).
De beweringen welke bewezen zijn onder (i), (ii) en (iii) vormen samen het bewijs van stelling 4.8.

Nauwkeurigheid voor predictor-corrector methoden

Stelling 4.9

Wanneer een P.C.-paar (4.23), (4.24) gegeven is, waarbij de predictor- en corrector-formule respectievelijk de orde van consistentie p en q bezitten, dan is de orde van dit P.C.-paar

$$\min(p+m,q),$$

waarin m het aantal correctieslagen aangeeft.

Bewijs

Laten de exacte waarden van y_{n-i} beschreven worden door $\tilde{y}_{n-i}$. We berekenen $\tilde{y}_n - y_n$ in het geval dat $\tilde{y}_{n-i} = y_{n-i}$ $(1 \leq i \leq k)$. Substitutie van $\tilde{y}_i$ en y_i in (4.20) en (4.21) levert

i) $$\tilde{y}_n - {}_0y_n = O(h^{p+1})$$

ii) $$\tilde{y}_n - {}_{(m+1)}y_n = \beta_0^* \, h(f(\tilde{y}_n) - f({}_my_n)) + O(h^{q+1}).$$

Wanneer we, zoals gebruikelijk, aannemen dat f aan een Lipschitz voorwaarde voldoet, laat men met volledige inductie eenvoudig zien dat geldt

$$\tilde{y}_n - {}_my_n = O(h^{p+m+1}) + O(h^{q+1}).$$

Stabiliteit voor predictor-corrector methoden

We beschouwen P.C.-methoden als een proces van de vorm (4.17) en we gaan na op welke wijze een verstoring in de vector $\vec{y}_n$ doorwerkt in de vector $\vec{y}_{n+1}$. We passen de methode weer toe op de modelvergelijking

$$y' = \delta y + f(x)$$

waarin δ een willekeurig complex getal is. Laat $\vec{y}_n^*$ de vector zijn die ontstaat door verstoring van $\vec{y}_n$ (bijvoorbeeld door afrondingsfouten).
We definieren nu

$$\vec{\varepsilon}_n = \vec{y}_n^* - \vec{y}_n$$

en we nemen aan dat bij de uitvoering van het proces geen nieuwe fouten optreden. Nu wordt $\vec{\varepsilon}_{n+1}$ berekend met (4.17)

$$
\begin{aligned}
{}_0\vec{\varepsilon}_{n+1} &= B\,\vec{\varepsilon}_n \\
{}_{(m+1)}\vec{\varepsilon}_{n+1} &= {}_m\vec{\varepsilon}_{n+1} + \vec{c}(h\delta\ {}_m\varepsilon_{n+1} - h\ {}_m\varepsilon'_{n+1}) \\
\vec{\varepsilon}_{n+1} &= M\vec{\varepsilon}_{n+1}.
\end{aligned}
$$

Dit kunnen we formuleren als

$$
\vec{\varepsilon}_{n+1} = (I + cp^T(h\delta))^M\, B\, \vec{\varepsilon}_n \tag{4.29}
$$

waarin $p^T(h\delta) = (h\delta,0,\ldots,0,-1,0,\ldots,0)$.
Voor stabiliteit van een P.C.-methode moet de spectraalradius van de matrix

$$
(I + cp^T(h\delta))^M\, B
$$

kleiner zijn dan 1. Het stabiliteitsgebied van de methode wordt begrensd door gedeelten van de kromme in het $h\delta$-vlak die bepaald wordt door

$$
\det((I+cp^T(h\delta))^M\, B - e^{i\phi}\, I) = 0 \qquad 0 \leq \phi \leq 2\pi.
$$

Dit is een gesloten kromme welke in het algemeen een aantal dubbelpunten zal bezitten.

<u>$P(EC)^M$ en $P(EC)^ME$ methoden</u>

De laatste correcties die in (4.17) worden uitgevoerd luiden

$$
\begin{aligned}
y_n &= {}_{(M-1)}y_n + \beta_0^* (hf({}_{(M-1}y_n) - h\ {}_{(M-1)}y'_n), \\
h\, y'_n &= h\ {}_{(M-1)}y'_n + (hf({}_{(M-1)}y_n) - h\ {}_{(M-1)}y'_n).
\end{aligned}
$$

Hierin wordt y_n bepaald met behulp van de laatst geëvalueerde functiewaarde $f({}_{(M-1)}y_n)$, welke waarde tevens als y'_n wordt gebruikt. Deze wijze van handelen wordt aangegeven met $P(EC)^M$: éénmaal wordt een predictie uitgevoerd en vervolgens worden M malen een functie-evaluatie en een correctieslag uitgevoerd. Een kleine wijziging kan hierop worden aangebracht door na de laatste

correctie nog eenmaal een functie-evaluatie uit te voeren, waarna uitsluitend y'_n gewijzigd wordt. Dit wordt aangegeven met $P(EC)^M E$. Door deze laatste functie-evaluatie wordt ook de stabiliteit beïnvloed. Het analogon van (4.29) luidt

$$\vec{\varepsilon}_{n+1} = (I + c^* p^T(h\delta))(I + cp^T(h\delta))^M B \vec{\varepsilon}_n,$$

waarin $c^* = (0,0,\ldots,0,1,0,\ldots,0)$ met een 1 op de (k+2)-de plaats. Hierdoor is het stabiliteitsgedrag van de $P(EC)^M E$ methoden vastgelegd.

Convergentieversnelling

Een noodzakelijke voorwaarde voor de bruikbaarheid van (4.17) is de convergentie van het iteratieve proces. De convergentievoorwaarde was reeds gevonden in (4.22). We merken op dat deze voorwaarde overeenkomt met de eis dat de matrix $I + cp^T(h\delta)$ uit formule (4.29) uitsluitend eigenwaarden binnen de eenheidscirkel heeft.

Een geschikte methode om het iteratieve proces te versnellen wordt gevonden in de Newton-Raphson- of de gewijzigde Newton-Raphson-methode voor het oplossen van (zie (4.17))

$$\vec{y}_n - {}_m\vec{y}_n = \vec{c}(hf(y_n) - h\, {}_m y'_n).$$

Bij het toepassen van deze methoden is het echter noodzakelijk dat men over de partiële afgeleide $\delta = \frac{\partial f}{\partial y}$ beschikt en dat men in iedere iteratieslag een lineaire vergelijking of stelsel vergelijkingen oplost.
Het iteratieve proces wordt dan beschreven door

$$(4.31) \qquad {}_{(m+1)}\vec{y}_n = {}_m\vec{y}_n + \vec{c}(1-h\delta\, \beta_0^*)^{-1} (hf({}_m y_n) - h\, {}_m y'_n).$$

De stabiliteit van dit nieuwe proces wordt berekend op dezelfde wijze als van proces (4.17). Het analogon van (4.29) luidt

$$(4.32) \qquad \vec{\varepsilon}_{n+1} = (I+(1-h\delta\, \beta_0^*)^{-1}\, cp^T(h\delta))^M B \vec{\varepsilon}_n .$$

Er geldt nu voor $I + (1-h\delta\,\beta_0^*)^{-1}\,cp^T(h\delta)$, in tegenstelling tot de overeenkomstige matrix in (4.29), dat alle eigenwaarden in absolute waarde kleiner dan of gelijk aan 1 zijn.

Equivalente beschrijving

Nordsieck heeft in 1962 een algorithme voor een lineaire meerstapsmethode gepubliceerd, die de informatie over het verleden niet in de vorm $\vec{y}_n$ gebruikt. Later heeft Gear (1967) er op gewezen dat het voor iedere lineaire meerstapsmethode mogelijk is een beschrijving te geven die niet de waarden y_{n-i} en $h\,y'_{n-i}$ zelf, maar een lineaire combinatie van y_{n-i} en $h\,y'_{n-i}$ gebruikt. Laat T een niet singuliere matrix zijn, dan worden $\vec{z}_n$ en ${}_m\vec{z}_n$ gedefinieerd door

$$\vec{z}_n = T\,\vec{y}_n \quad \text{en} \quad {}_m\vec{z}_n = T\,{}_m\vec{y}_n \,.$$

Laat T bovendien zodanig zijn dat

$$z_n = y_n \quad \text{en} \quad h\,z'_n = h\,y'_n$$

dan beschrijft

$$\begin{aligned} {}_0\vec{z}_n &= TBT^{-1}\,\vec{z}_{n-1} \\ {}_{(m+1)}\vec{z}_n &= {}_m\vec{z}_n + T\,\vec{c}(hf({}_mz_n) - h\,{}_mz'_n) \\ \vec{z}_n &= M\,\vec{z}_n \end{aligned} \tag{4.33}$$

een methode die equivalent is met (4.17). Hoewel de analytische eigenschappen van equivalente methoden gelijk zijn, is het mogelijk dat de implementatie in het ene geval aantrekkelijker is dan in het andere geval.

De bovengenoemde methode van Nordsieck (1962) is equivalent met de Adams-Moulton methode. De vector

$$\vec{y}_n = (y_n, hy'_n, hy'_{n-1}, \ldots, hy'_{n-k})$$

van de Adams-Moulton methode is hierbij getransformeerd tot de vector

$$\vec{z}_n = (y_n, hy'_n, h^2y''_n, \ldots, h^{k+1} y_n^{(k+1)}/(k+1)!).$$

Deze transformatie heeft het voordeel dat betrekkelijk eenvoudig op een andere (vaste) staplengte overgeschakeld kan worden.

Lineaire meerstapsmethoden met variabele staplengte

We zullen ons bij de behandeling van lineaire meerstapsmethoden met variabele staplengte beperken tot de formules van het Adams-Bashforth-, Adams-Moulton- en Curtiss-Hirschfelder-type. Deze methoden onderscheiden zich daardoor, dat met een minimale hoeveelheid informatie uit het verleden (d.w.z. met een korte vector $\vec{y}_n$) een maximale orde van consistentie verkregen wordt. De Adams-Moulton-methoden zijn sterk stabiele k-stapsmethoden met de maximaal bereikbare orde k + 1. De methoden van het type zoals voorgesteld door Curtiss en Hirschfelder (1952), zijn instabiel voor orde $p > 6$ (Mitchell en Craggs, 1953) en zijn (daarom?) tot 1968 praktisch niet gebruikt. Gear (1968) heeft echter aangetoond dat deze methoden, voor orde $p \leq 6$, "stiffly-stable" zijn, waardoor ze bijzonder geschikt zijn voor het oplossen van stijve differentiaalvergelijkingen.

De genoemde methoden laten zich eenvoudig karakteriseren door de keuze van sommige coëfficiënten in de algemene formule voor lineaire meerstapsmethoden (4.2). We geven deze coëfficiënten in tabel 4.2

Tabel 4.2

methode	coëfficiënten	orde
Adams-Bashforth	$\alpha_i = 0$ $(2 \leq i \leq k)$ $\beta_0 = 0$	k
Adams-Moulton	$\alpha_i = 0$ $(2 \leq i \leq k)$	k + 1
Curtiss-Hirschfelder	$\beta_i = 0$ $(1 \leq i \leq k)$	k

De niet in tabel 4.2 genoemde coëfficiënten worden allen bepaald door de (maximale) orde voor deze k-stapsmethoden.

De Adams-Bashforth- en Adams-Moulton-methoden kunnen beschouwd worden als methoden waarbij de afgeleide y'(x) benaderd wordt door een (k-1)ste resp. k-de graads polynoom h(x) dat bepaald wordt door
$y'_i = h(x_i)$ $n-k \leq i < n$ (resp. $n-k \leq i \leq n$).

De Curtiss-Hirschfelder-methoden benaderen de oplossing y(x) met een k-de graads polynoom h(x) door de punten y_i ($n-k \leq i < n$); de vrijheid die overblijft bij de constructie van dit polynoom wordt gebruikt om aan de relatie $y'_n = h'(x_n) = f(x_n,y_n)$ te voldoen.

Benadering van een functie door een polynoom

Men kan een polynoom h(x), waarvan de waarde voor een aantal steunpunten vastligt, beschrijven door gebruik te maken van Newton's interpolatieformule met differentiequotiënten. Om deze beschrijving op eenvoudige wijze te kunnen geven gebruiken we een vectornotatie.

Definities. Laten x_i ($n-k \leq i \leq n$) de steunpunten voor een k-stapsmethode zijn; we definiëren dan de rij-vectoren H(x) en G(x) als volgt:

$$H(x) = (1,(x-x_{n-1}),(x-x_{n-1})^2,\ldots,(x-x_{n-1})^k),$$

$$G(x) = (1,(x-x_{n-1}),(x-x_{n-1})(x-x_{n-2}),\ldots,(x-x_{n-1})(x-x_{n-2})\ldots(x-x_{n-k})).$$

Een k-de graads polynoom h(x) waarvan de waarden bekend zijn voor de steunpunten x_i ($n-k \leq i \leq n$) laat zich nu schrijven als produkt van G(x) en de kolomvector van differentiequotiënten a:

$$h(x) = G(x).a \tag{4.34}$$

met $a = (h[x_{n-1}],h[x_{n-1},x_{n-2}],\ldots,h[x_{n-1},\ldots,x_{n-k}],h[x_{n-1},\ldots,x_{n-k},x_n])^T$
waarin

$$\begin{cases} h[x_i] = h(x_i) & \text{en} \\ h[x_i,\ldots,x_j] = (h[x_i,\ldots,x_{j-1}]-h[x_{i+1},\ldots,x_j]) \;/\; (x_i-x_j). \end{cases} \tag{4.35}$$

Het polynoom h(x) kan ook geschreven worden als produkt van H(x) met een kolomvector p

$$h(x) = H(x).p \tag{4.36}$$

met $p = (h(x_{n-1}),h'(x_{n-1}),h''(x_{n-1})/2!,\ldots,h^{(k)}(x_{n-1})/k!)$. Deze vector p heeft als componenten de Taylor-coëfficiënten in het punt x_{n-1}.

Wanneer we in de volgende paragraaf de formules (4.34) en (4.36) toepassen, is in eerste instantie de functiewaarde $h(x_n)$ onbekend. In dat geval is, van de vector a, alleen de laatste component niet bepaald. Van de vector p daarentegen kan, behalve de eerste, geen enkele component berekend worden.

Door het definitie-gebied van de functie h(x) te beperken tot de steunpunten x_i $(n-k\leq i\leq n)$ kunnen we een vector h definiëren:

$$h \triangleq (h(x_{n-1}),h(x_{n-2}),\ldots,h(x_{n-k}),h(x_n))^T.$$

Op dezelfde wijze definiëren we de matrices H en G door uitsluitend de functiewaarden H(x) en G(x) voor de argumenten x_i $(n-k\leq i\leq n)$ te beschouwen.

Voorbeeld voor k = 3

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & (x_{n-2}-x_{n-1}) & 0 & 0 \\ 1 & (x_{n-3}-x_{n-1}) & (x_{n-3}-x_{n-1})(x_{n-3}-x_{n-2}) & 0 \\ 1 & (x_n-x_{n-1}) & (x_n-x_{n-1})(x_n-x_{n-2}) & (x_n-x_{n-1})(x_n-x_{n-2})(x_n-x_{n-3}) \end{pmatrix}$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & (x_{n-2}-x_{n-1}) & (x_{n-2}-x_{n-1})^2 & (x_{n-2}-x_{n-1})^3 \\ 1 & (x_{n-3}-x_{n-1}) & (x_{n-3}-x_{n-1})^2 & (x_{n-3}-x_{n-1})^3 \\ 1 & (x_n-x_{n-1}) & (x_n-x_{n-1})^2 & (x_n-x_{n-1})^3 \end{pmatrix}$$

Direct hieruit volgt nu het gediscretiseerde analogon van (4.34) en (4.36)

$$Ga = h = Hp.$$

De vector p kan nu uitgedrukt worden in a, G en H

$$p = H^{-1}Ga. \tag{4.37}$$

We definiëren de matrix A door

$$A \stackrel{d}{=} H^{-1}G. \tag{4.38}$$

Deze matrix heeft de eigenschap dat

$$h(x) = G(x)a = H(x)Aa. \tag{4.39}$$

Een andere eigenschap maakt het mogelijk de matrix A eenvoudig te construeren. Deze eigenschap luidt:

De matrix A is een bovendriehoeksmatrix waarvan de elementen worden bepaald door

$A_{00} = 1, \quad A_{0j} = 0 \qquad (j \neq 0),$

$A_{ij} = s_{j-1} A_{i,j-1} + A_{i-1,j-1} \qquad (0 < i \leq j);$

waarin $s_j = x_{n-1} - x_{n-j-1}$.

Voor het bewijs van deze eigenschap zij verwezen naar Hemker (1971).

Voorbeeld voor k = 3

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & s_1 & s_1 s_2 \\ 0 & 0 & 1 & s_1+s_2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We merken op dat in de matrix A uitsluitend de afstanden tussen de steunpunten x_i $(n-k \leq i < n-1)$ voorkomen.

De formules van Curtiss-Hirschfelder voor variabele staplengte

Voor het uitvoeren van een stap van het integratieproces volgens de methode van Curtiss en Hirschfelder staan de waarden $y_{n-k}, \ldots, y_{n-1}$ ter beschikking en dient y_n te voldoen aan

$$h'(x_n) = f(x_n, y_n), \tag{4.40}$$

waarbij $h(x)$ het k-de graads polynoom is, dat bepaald wordt door $h(x_i) = y_i$ $(n-k \leq i \leq n)$. Behalve het laatste differentiequotiënt $h[x_{n-1}, \ldots, x_{n-k}, x_n]$ kunnen de componenten van de vector a eenvoudig berekend worden (zie (4.35)). We splitsen de vector a daarom in een bekend deel

$$a_{n-1} = (y_{n-1}, h[x_{n-1}, x_{n-2}], \ldots, h[x_{n-1}, \ldots, x_{n-k}], 0)^T$$

en een onbekend deel

$$a - a_{n-1} = (0, \ldots, 0, 1)^T \, h[x_{n-1}, \ldots, x_{n-k}, x_n]$$

zodat uit (4.39) volgt

$$h(x) = H(x)A\, a_{n-1} + H(x)\, A \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} h[x_{n-1}, \ldots, x_{n-k}, x_n]. \tag{4.41}$$

We definiëren p_{n-1}, de vector van Taylor-coëfficiënten in het punt x_{n-1} van het (k-1)de-graads polynoom door $y_{n-1}, \ldots, y_{n-k}$, door

$$p_{n-1} \triangleq A\, a_{n-1}, \tag{4.42}$$

en we geven met A_n de laatste kolom van A aan:

$$A_n \triangleq A \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{4.43}$$

We mogen (4.41) nu schrijven als

$$h(x) = H(x)\, p_{n-1} + H(x)\, A_n\, h[x_{n-1},\ldots,x_{n-k},x_n]. \tag{4.44}$$

We noteren de differentiaaloperator d/dx met D.

Differentiëren van (4.44) levert

$$D^j h(x) = D^j H(x)\, p_{n-1} + D^j H(x)\, A_n\, h[x_{n-1},\ldots,x_{n-k},x_n] \qquad (j=0,1,2,\ldots). \tag{4.45}$$

We gebruiken de eerste term van het rechterlid voor het berekenen van een "voorspelling" van $y_n^{(j)}$:

$$D^j h(x_n)^{pred} \triangleq D^j H(x_n)\, p_{n-1}, \tag{4.46}$$

waarbij later een "correctie" $\Delta D^j h(x_n)$ zal worden opgeteld:

$$\Delta D^j h(x_n) \triangleq D^j H(x_n)\, A_n\, h[x_{n-1},\ldots,x_{n-k},x_n]. \tag{4.47}$$

Het differentiequotiënt $h[x_{n-1},\ldots,x_{n-k},x_n]$ is onbekend; door herhaald toepassen van formule (4.47) volgt echter

$$\Delta D^j h(x_n) = \frac{D^j H(x_n) A_n}{D\, H(x_n) A_n}\, \Delta D\, h(x_n), \tag{4.48}$$

en uit (4.45) en (4.40) volgt

$$\begin{aligned} \Delta D\, h(x_n) &= D\, h(x_n) - D\, h(x_n)^{pred} \\ &= f(x_n,y_n) - D\, h(x_n)^{pred}, \end{aligned} \tag{4.49}$$

zodat we nu een iteratieschema kunnen opstellen om y_n op te lossen uit (4.49). Dit iteratieschema bestaat, evenals (4.17), uit een predictie, een aantal correcties en de definitie van de berekende waarde y_n.

$${}_0y_n^{(j)} = D^j H(x_n)\, p_{n-1} \qquad (0 \le j \le k) \tag{4.50}$$

$$(4.51)\qquad \begin{cases} {}^{(m+1)}y_n = {}^{m}y_n + \dfrac{H(x_n)A_n}{DH(x_n)A_n}\,(f(x_n,{}^{m}y_n)-{}^{m}y_n') \\ {}^{(m+1)}y_n' = {}^{m}y_n' + \qquad\qquad (f(x_n,{}^{m}y_n)-{}^{m}y_n') \end{cases}$$

$$(4.52)\qquad \begin{cases} y_n = {}^{M}y_n \\ y_n' = {}^{M}y_n \\ y_n^{(j)} = {}^{0}y_n + \dfrac{D^jH(x_n)A_n}{D\,H(x_n)A_n}\,(y_n'-{}^{0}y_n') \qquad j = 2,\ 3,\ \ldots,\ k \\ p_n = (y_n, y_n', \ldots, y_n^{(k)}/k!). \end{cases}$$

De quotiënten $\dfrac{D^jH(x_n)A_n}{D\,H(x_n)A_n}$ zijn functies van de waarden x_i $(n-k\leq i\leq n)$. We merken op dat voor de berekening van A_n alleen de waarden $s_j = x_{n-1} - x_{n-j-1}$ nodig zijn, terwijl voor de berekening van $D^jH(x_n)$ alleen de laatste staplengte $x_n - x_{n-1}$ nodig is. Wanneer we aannemen dat alle staplengten van dezelfde orde van grootte zijn, volgt na enige berekening

$$D^jH(x_n)A_n = O(h^{k-j}).$$

Voorbeeld

Als illustratie voeren we de berekening van de quotiënten $\dfrac{D^jH(x_n)A_n}{D\,H(x_n)A_n}$ uit voor $k = 3$, $s_j = jh$ en $x_n - x_{n-1} = h$.
A_n is de kolomvector $(A_{ij})_{j=3}$:

$$A_n = (0, s_1 s_2, s_1+s_2, 1)^T = (0, 2h^2, 3h, 1)^T.$$

Uit definitie van $H(x)$ volgt

$$\begin{aligned} H(x_n) &= (1, h, h^2, h^3) \quad \text{zodat} \quad & H(x_n)A_n &= 6h^3, \\ D\,H(x_n) &= (0, 1, 2h, 3h^2) & D\,H(x_n)A_n &= 11h^2, \\ D^2H(x_n) &= (0, 0, 2, 6h) & D^2H(x_n)A_n &= 12h, \\ D^3H(x_n) &= (0, 0, 0, 6) & D^3H(x_n)A_n &= 6. \end{aligned}$$

De quotiënten $\dfrac{D^jH(x_n)A_n}{D\,H(x_n)A_n}$ zijn dus voor $j = 0,\ 1,\ 2,\ 3$ respectievelijk

$\frac{6}{11}h$, 1, $\frac{12}{11}h^{-1}$ en $\frac{6}{11}h^{-2}$.

Uit (4.51) volgt direkt de voorwaarde voor convergentie:

$$\left| \frac{H(x_n)A_n}{DH(x_n)A_n} \frac{\partial f}{\partial y} \right| < 1.$$

Voor het oplossen van stijve differentiaalvergelijkingen is dit iteratieproces ongeschikt omdat grote waarden van $|\partial f/\partial y|$ een kleine waarde van $\frac{H(x_n)A_n}{DH(x_n)A_n}$ en daarmee een kleine staplengte noodzakelijk maken. We versnellen het proces daarom met de (gewijzigde) Newton-Raphson methode. De iteratie (4.51) wordt dan vervangen door

$$(4.53) \quad \begin{cases} {}^{(m+1)}y_n = {}^{m}y_n + \frac{H(x_n)A_n}{DH(x_n)A_n} \left(I - \frac{H(x_n)A_n}{DH(x_n)A_n} \frac{\partial f}{\partial y}\right)^{-1} \left(f(x_n, {}^{m}y_n) - {}^{m}y_n'\right) \\ {}^{(m+1)}y_n' = {}^{m}y_n' + \left(I - \frac{H(x_n)A_n}{DH(x_n)A_n} \frac{\partial f}{\partial y}\right)^{-1} \left(f(x_n, {}^{m}y_n) - {}^{m}y_n'\right). \end{cases}$$

Dit kan men direkt afleiden uit

$${}^{(m+1)}y_n' - {}^{m}y_n' = f(x_m, {}^{m}y_n) - {}^{m}y_n' + \left({}^{(m+1)}y_n - {}^{m}y_n\right) \frac{\partial f}{\partial y}$$

$$= f(x_n, {}^{m}y_n) - {}^{m}y_n' + \frac{H(x_n)A_n}{DH(x_n)A_n} \frac{\partial f}{\partial y} \left({}^{(m+1)}y_n' - {}^{m}y_n'\right).$$

De Adams-Bashforth-Moulton formules voor variabele staplengte

Het formularium voor de Adams-Bashforth-Moulton methoden verschilt slechts weinig van dat van de Curtiss-Hirschfelder methoden. Hier wordt niet $y(x)$, maar $y'(x)$ benaderd door een polynoom $h(x)$ zodat $h(x_i) = y_i'$ $n-k \leq i \leq n$. Met de volgende definitie van de vector der differentiequotiënten:

$$a_{n-1} = (y_{n-1}', h[x_{n-1}, x_{n-2}], \ldots, h[x_{n-1}, \ldots, x_{n-k}], 0)^T,$$

blijven (4.41) t/m (4.48) onveranderd geldig. De formules (4.45) t/m (4.48) gelden tevens voor $j = -1$ wanneer de integratieoperator D^{-1} gebruikt wordt

voor integratie van x_{n-1} tot x_n en wanneer bovendien de juiste integratieconstante gekozen wordt. Zo wordt de predictie van $y_n^{(j)}$ gegeven door (vgl. (4.46))

$$(4.54)\qquad \begin{cases} y_n{}^{\text{pred}} = y_{n-1} + D^{-1}\ H(x_n)p_{n-1} \\ y_n^{(j)\text{pred}} = \qquad D^{j-1}\ H(x_n)p_{n-1} \qquad (j>0). \end{cases}$$

Deze predictor is juist de Adams-Bashforth formule voor variabele staplengte.
Uit (4.40) volgt nu, anders dan bij de Curtiss-Hirschfelder-methoden, - vergelijk (4.49) -, dat

$$(4.55)\qquad \begin{aligned} \Delta h(x_n) &= h(x_n) - h(x_n)^{\text{pred}} \\ &= f(x_n,y_n) - h(x_n)^{\text{pred}}, \end{aligned}$$

zodat hier het iteratieproces luidt:

$$(4.56)\qquad \begin{cases} {}_0y_n = y_{n-1} + D^{-1}\ H(x_n)p_{n-1} \\ {}_0y_n^{(j)} = \qquad D^{j-1}\ H(x_n)p_{n-1} \qquad (j>0) \end{cases}$$

$$(4.57)\qquad \begin{cases} {}_{(m+1)}y_n = {}_my_n + \dfrac{D^{-1}H(x_n)A_n}{H(x_n)A_n}\ (f(x_n,{}_my_n)-{}_my_n') \\ {}_{(m+1)}y_n' = {}_my_n' + \qquad\qquad (f(x_n,{}_my_n)-{}_my_n') \end{cases}$$

$$(4.58)\qquad \begin{cases} y_n = {}_My_n \\ y_n' = {}_My_n' \\ y_n^{(j)} = {}_0y_n^{(j)} + \dfrac{D^{j-1}H(x_n)A_n}{H(x_n)A_n}\ (y_n'-{}_0y_n') \qquad (j>1) \\ p_n = (y_n',y_n'',\ldots,y_n^{(k+1)}/k!). \end{cases}$$

Ook hier kan de convergentie versneld worden met de (gewijzigde) Newton-Raphson methode; (4.57) wordt dan

$$(4.59)\begin{cases} {}^{(m+1)}y_n = {}^{m}y_n + \dfrac{D^{-1}H(x_n)A_n}{H(x_n)A_n}\left(I - \dfrac{D^{-1}H(x_n)A_n}{H(x_n)A_n}\dfrac{\partial f}{\partial y}\right)^{-1}\left(f(x_n, {}^{m}y_n) - {}^{m}y_n'\right) \\ {}^{(m+1)}y_n' = {}^{m}y_n' + \left(I - \dfrac{D^{-1}H(x_n)A_n}{H(x_n)A_n}\dfrac{\partial f}{\partial y}\right)^{-1}\left(f(x_n, {}^{m}y_n) - {}^{m}y_n'\right). \end{cases}$$

Stabiliteit en nauwkeurigheid voor de processen (4.50) en (4.56)

De analyse van de stabiliteit verloopt voor methoden met variabele staplengte op dezelfde wijze als voor de methoden met vaste staplengte, zoals die gegeven werd naar aanleiding van proces (4.17). De resultaten geven geen essentieel nieuwe gezichtspunten. Voor een uitgebreide behandeling en voor afbeeldingen van een aantal stabiliteitsgebieden zij verwezen naar Hemker (1971).

We merken op dat bij een iteratieschema zoals gegeven in (4.50), (4.51) en (4.52) of in (4.56), (4.57) en (4.58) de orde van consistentie voor de predictorformule één lager is dan voor de correctorformule. Zoals is aangetoond in stelling 4.9, is dan slechts één iteratieslag voldoende om de orde van consistentie van het proces gelijk te maken aan die van de correctorformule. Wanneer echter een beperkt aantal iteratieslagen wordt uitgevoerd, is de keuze van deze predictorformule gunstig met betrekking tot de stabiliteit. Als illustratie geven we in figuur 4.1 een aantal stabiliteitsgebieden voor Adams-Bashforth-Moulton-PECE-methoden en voor Curtiss-Hirschfelder-PEC-methoden. Telkens is in het complexe $h\delta$-vlak het stabiliteitsgebied getekend van een predictor-corrector-paar met gelijke orde en van een predictor-corrector-paar waarbij de orde van de predictor één lager is dan van de corrector. Telkens blijkt het stabiliteitsgebied voor de laatstgenoemde combinatie groter te zijn dan voor de eerstgenoemde.

Figuur 4.1

Stabiliteitsgebieden voor predictor-corrector methoden

Linkerzijde: Adams-Bashforth-Moulton-PECE-methoden.

Rechterzijde: Curtiss-Hirschfelder-PEC-methoden.

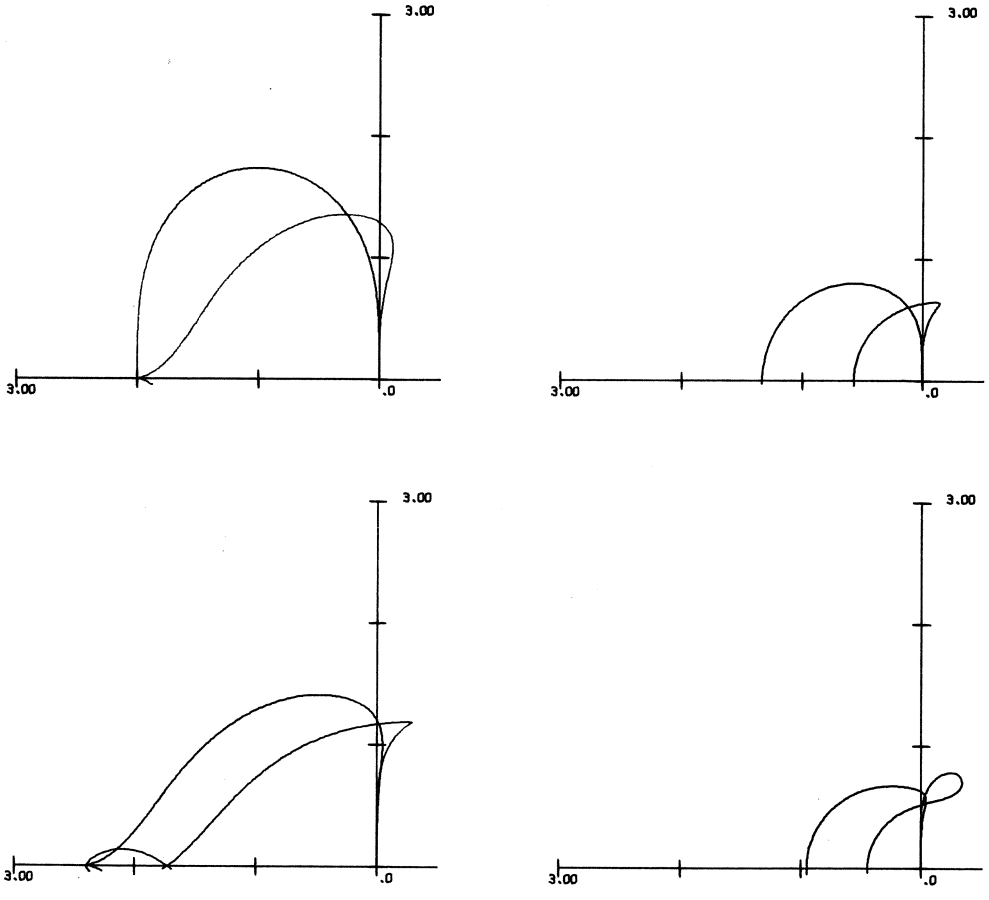
Van boven naar beneden: orde van consistentie

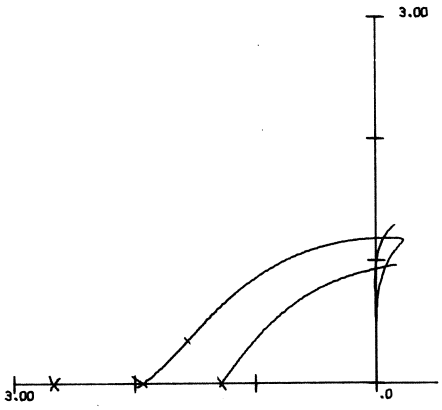
2, 3, 4, 5 en 6.

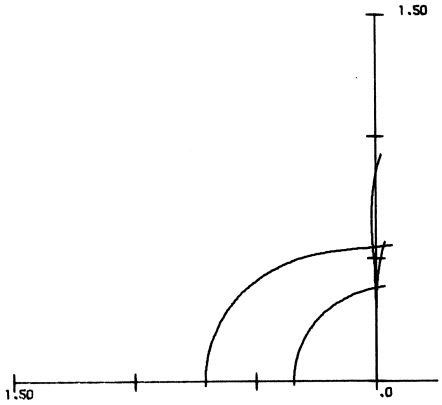
In elk figuur:

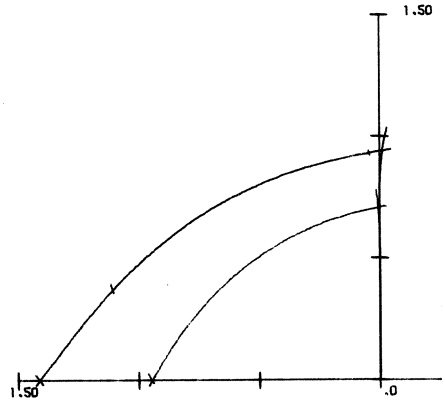
het kleine gebied: predictor en corrector van gelijke orde;

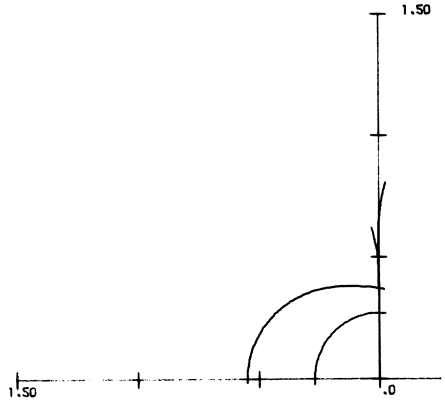
het grote gebied: predictor één orde lager.

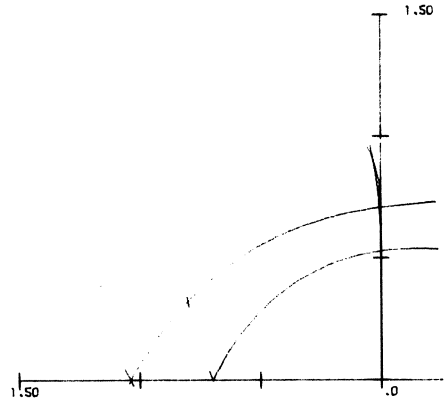


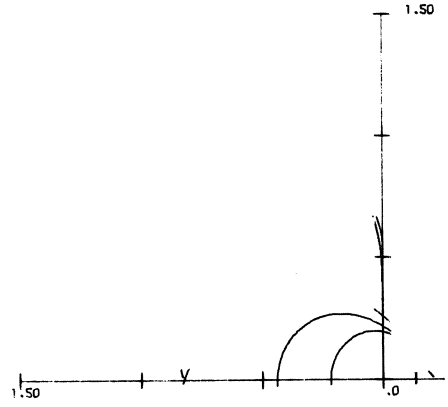

3.00
3.00
.0


1.50
1.50
.0


1.50
1.50
.0


1.50
1.50
.0


1.50
1.50
.0


1.50
1.50
.0

Literatuur

Curtiss, C.F. and Hirschfelder, J.O.
Integration of stiff equations.
Proc. Nat. Acad. Sci. U.S. 38 (1952) 235.

Dahlquist, G.
Convergence and stability in the numerical integration of ordinary differential equations.
Math. Scand. 4 (1956) 33.

Dahlquist, G.
A special stability problem for linear multistep methods.
BIT 3 (1963) 27.

Gear, C.W.
The numerical integration of ordinary differential equations.
Math. Comp. 21 (1967) 146.

Gear, C.W.
The automatic integration of stiff ordinary differential equations.
Proc. IFIP Congr. 1968 p. 187.

Gear, C.W.
Numerical initial value problems in ordinary differential equations.
Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.

Hemker, P.W.
Lineaire meerstapsmethoden met variabele staplengte.
Rapport NR 15/71, Mathematisch Centrum, Amsterdam.

Henrici, P.
Discrete variable methods in ordinary differential equations.
John Wiley and Sons, New York, 1962.

Mitchell, A.R. and Craggs, J.W.
Stability of difference relations in the solution of ordinary differential equations.
Math. Comp. 7 (1953) 127.

Nordsieck, A.

On numerical integration of ordinary differential equations.
Math. Comp. 16 (1962) 22.

Widlund, O.B.

A note on unconditionally stable linear multistep methods.
BIT 7 (1967) 65.