

COLLOQUIUM STIJVE DIFFERENTIAALVERGELIJKINGEN

DEEL 2

DOOR P.A.BEENTJES, K.DEKKER, H.C.HEMKER,

S.P.N. VAN KAMPEN, G.M.WILLEMS

Inhoud

Voorwoord

5. Parameterbepaling in de biomathematica	120
G.M. Willems, Math. Centrum, Amsterdam	
6. Foutformules en stapkeuzestrategieën	138
P.A. Beentjes, Math. Centrum, Amsterdam	
7. Exponentieel aangepaste methoden	153
K. Dekker, Math. Centrum, Amsterdam	
8. Semi Runge-Kutta methoden	175
S.P.N. van Kampen, Math. Centrum, Amsterdam	
9. Stijve differentiaalvergelijkingen in biochemische modellen	197
H.C. Hemker, Rijks Universiteit Leiden	

Voorwoord

In het colloquium 'stijve differentiaalvergelijkingen', dat van september 1971 tot december 1972 aan het Mathematisch Centrum gehouden werd, is een overzicht gegeven van recent onderzoek dat verricht werd op het gebied van het numeriek oplossen van beginwaardeproblemen.

In het eerste deel van de syllabus van dit colloquium werd, na een inleiding en een historisch overzicht, een behandeling gegeven van eenstaps- en meerstapsmethoden. In dit tweede deel van de syllabus wordt de nadruk gelegd op een aantal speciale onderwerpen zoals stapkeuze-strategieën, exponentieel aangepaste methoden en gegeneraliseerde Runge-Kutta methoden. Ook komt in dit tweede gedeelte een interessant toepassingsgebied naar voren: de analyse van biochemische modellen. Door de nieuwe numerieke technieken is het mogelijk geworden efficiënter biochemische processen te simuleren en - aan de hand van experimenteel verkregen gegevens - onbekende modelparameters te berekenen.

Het derde gedeelte van de syllabus zal weer geheel gewijd zijn aan numerieke methoden. Daarin zullen enige theoretische aspecten en de resultaten van een vergelijkend onderzoek behandeld worden.

P.W. Hemker.

5. Parameterbepaling in de biomathematica

In de biomathematica komt het nogal eens voor, dat men redelijkerwijs kan aannemen of zelfs kan bewijzen, dat het verloop in de tijd van de toestandsvariabele $x(t)$ van een bepaald systeem beschreven wordt door het beginwaardeprobleem

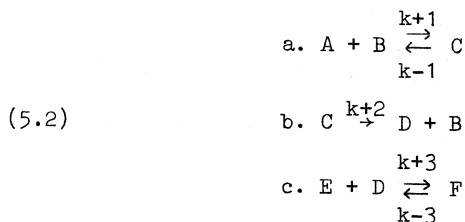
$$(5.1) \quad \dot{x}(t) = f(x(t), t, \text{par}) \quad , \quad x(t_0) = x_0 \quad ,$$

waarbij de toestandsvariabele $x(t)$ een continu differentieerbaar van de tijd afhankelijke vector met n componenten is, $f : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ een gegeven functie van x , t en par , $\text{par} \in \mathbb{R}^m$, is en waarbij par een nog te bepalen bij het systeem behorende vector is. Aangezien het veelal onmogelijk is de parameters par , die de juiste beschrijving van het systeem opleveren, op directe wijze te bepalen, ontstaat nu het probleem om deze uit observaties van het tijdsverloop van de toestandsvariabele $x(t)$ af te leiden. Dit alles leidt tot de vraag om bij gegeven f uit (5.1) en bij voor verscheidene tijden $t \in T$ gegeven waarden van observaties $\text{obs}(t)$ van $x(t)$ die met een zekere fout belast zijn (dat wil zeggen $\text{obs}(t) = x(t) + E(t)$) een schatting te maken van de waarde van de parametervector par , die bij $x(t)$ hoort.

In deze voordracht zal een methode om dergelijke schattingen te maken besproken worden. Voorts zal deze methode op een concreet, uit de biochemie afkomstig, probleem toegepast worden. Hierbij zullen we in de gelegenheid zijn om te laten zien, dat in dit geval enige van de tot dusver in het colloquium behandelde numerieke integratie methoden met vrucht gebruikt kunnen worden.

Voorbeelden

Veronderstel, dat men op de één of andere manier heeft ontdekt dat, bij het ontstaan van het eindproduct F , chemische reacties met de stoffen A , B , C , D en E betrokken zijn en veronderstel voorts, dat men plausibel maken kan, dat dit alles volgens het volgende reactieschema verloopt



Wanneer men in dit stadium van het onderzoek is gekomen, zal men trachten elk van de afzonderlijke reacties te bestuderen en zo de reactieconstanten k_{+1} , k_{+2} , k_{+3} te bepalen. Veelal blijkt dit nu onmogelijk te zijn en zal het bovenstaande, namelijk het bepalen van de reactieconstanten en het verifiëren, moeten gebeuren aan de hand van het gedrag van het gehele systeem. En dit nu is een concreet geval van het in de inleiding genoemde probleem:

Het is namelijk wel goed mogelijk om met behulp van de wet van de massawerking differentiaalvergelijkingen, die het verloop in de tijd van de concentraties A, B, C, D, E en F* beschrijven, voor nog nader te bepalen waarden van de reactie constanten op te stellen:

$$\begin{array}{ll}
 \text{a. } \dot{A} = -k_{+1}AB + k_{-1}C, & A(t_0) = A_0 \\
 \text{b. } \dot{B} = -k_{+1}AB + k_{-1}C + k_{+2}C, & B(t_0) = B_0 \\
 \text{c. } \dot{C} = k_{+1}AB - k_{-1}C - k_{+2}C, & C(t_0) = C_0 \\
 (5.3) & \\
 \text{d. } \dot{D} = k_{+2}C + k_{-3}F - k_{+3}ED, & D(t_0) = D_0 \\
 \text{e. } \dot{E} = k_{-3}F - k_{+3}ED, & E(t_0) = E_0 \\
 \text{f. } \dot{F} = k_{+3}ED - k_{-3}F, & F(t_0) = F_0
 \end{array}$$

De vergelijkingen (5.3 a) en (5.3 b), (5.3 c) en (5.3 d), en (5.3 a), (5.3 c), (5.3 d) en (5.3 f) zijn afhankelijk. Het stelsel (5.3) kan door eliminatie van deze afhankelijkheden gereduceerd worden tot het stelsel:

*Bij een stof A wordt de hoofdletter A tevens gebruikt om de concentratie aan te duiden.

$$\begin{aligned}
 (5.4) \quad & \text{a. } \dot{A} = -k_{+1}AB + k_{-1}C, A(t_0) = A_0, \\
 & \text{b. } \dot{C} = k_{+1}AB - (k_{+2} + k_{-1})C, C(t_0) = C_0, \\
 & \text{c. } \dot{D} = k_{+2}C + k_{-3}F - k_{+3}ED, D(t_0) = D_0,
 \end{aligned}$$

waarbij voor B, E en F de relaties

$$\begin{aligned}
 (5.4') \quad & \text{a. } B = B_0 + C_0 - C, \\
 & \text{b. } E = E_0 + F_0 - F, \\
 & \text{c. } F = F_0 + A_0 + C_0 + D_0 - A - C - D,
 \end{aligned}$$

gelden.

De vraag om uit de gegeven observaties op verschillende tijden $t, t \in T$, van de toestandsvariabelen (A, C, D) van het systeem (4.2) de reactieconstanten k_{+1}, k_{+2}, k_{+3} te bepalen, is een concreet geval van het in de inleiding van deze voordracht gestelde probleem. Betreffende dit eerste voorbeeld valt nog op te merken, dat het in (5.2) respectievelijk (5.4) voorkomende deelsysteem, gevormd door (5.2 a) en (5.2 b) respectievelijk (5.4 a) en (5.4 b), hetzelfde is als het in de inleidende voordracht door P.W. Hemker besproken Michaelis-Menten systeem.

Op pagina 12 van de syllabus van deze inleidende voordracht is een gereduceerde vorm afgeleid voor het stelsel differentiaalvergelijkingen (1.6), dat bij de Michaelis-Menten vergelijkingen -in het vervolg om voor de hand liggende redenen ESCEP genoemd- het verloop in de tijd van het systeem beschrijft.

Het gereduceerde stelsel had de volgende gedaante:

$$\begin{aligned}
 (5.5) \quad & \dot{s} = (c-1)s + \text{par}_1 c, s(0) = 1 \\
 & \dot{c} = \text{par}_2 ((1-c)s - \text{par}_3 c), c(0) = 0
 \end{aligned}$$

In deze voordracht zullen wij het schatten van de parameters $\text{par}_1, \text{par}_2$ en par_3 uit (5.5) met behulp van kennis van (met fout belaste) waarden van s en c op verschillende tijdstippen t als modelprobleem gebruiken.

Het residu

Allereerst formuleren we nogmaals het in de inleiding genoemde probleem.

Zij gegeven:

- (5.6)
- a. een verzameling tijden T , waarop de observaties plaats gevonden hebben.
 - b. de observaties $\text{obs}(t)$ van $x(t)$ (eventueel van een aantal componenten van $x(t)$; aangezien dit voor de hier gevolgde methode van parameterschatting slechts een eenvoudige wijziging in de procedure impliceert, wordt aan dit geval in de rest van het betoog voorbijgegaan) op alle tijden $t \in T$.
 - c. de functie f uit (5.1).
 - d. het feit dat $x(t)$ de oplossing van het beginwaardeprobleem (5.1) is, bij beginwaarde $x(t_0) = x_0$ en bij onbekend veronderstelde waarde par_x van de parametervector par .

Dan wordt gevraagd een schatting te maken van par_x .

We gebruiken in het vervolg de notatie $x(t, \text{par})$ om de oplossing van het beginwaardeprobleem (5.1) met beginwaarde $x(t_0, \text{par}) = x_0$ en met rechterlid f uit (4.6), waarin voor de parameter par gesubstitueerd is, aan te duiden.

Aangezien we geen kennis over de structuur van de fout in de observaties veronderstellen, zullen we elke parameter par zodanig, dat $x(t, \text{par})$ binnen de fout met $\text{obs}(t)$ overeenstemt, als een goede schatting van par_x moeten beschouwen. Dit wil zeggen dat we gezien onze kennis (5.6) omtrent par_x geen reden hebben om te veronderstellen, dat deze waarde van de parameter niet de juiste (par_x) is. Bovenstaande overwegingen leiden tot de volgende tamelijk gangbare procedure voor het schatten van par_x (zie bijv. Lee [1968]):

- (5.7)
- a. Eerst wordt een norm $|| \cdot ||_T$ gekozen op de ruimte $C(T, \mathbb{R}^n)$ van continue functies, die T in $\mathbb{R}^n$ afbeelden.
 - b. Dan wordt de schatting par_s van par_x zo gekozen, dat

$$\text{Res}_{||\cdot||_T}(\text{par}_s) = \inf_{\text{par}} \text{Res}(\text{par}),$$

(5.7) waarbij het residu $\text{Res}_{||\cdot||_T}(\text{par})$ gedefinieerd is door

$$\text{Res}_{||\cdot||_T}(\text{par}) = ||x(\cdot, \text{par}) - \text{obs}(\cdot)||_T;$$

We schatten par_x dus door de parameter, die de beste benadering $x(t, \text{par})$ van $\text{obs}(t)$ oplevert.

Het toepassen van de hierboven genoemde procedure om de parameters par_x te schatten zal in deze voordracht verder besproken worden. Wij zullen dan voor $||\cdot||_T$ steeds de norm

$$||f|| = \left(\sum_{t \in T} ||f(t)||_2^2 \right)^{\frac{1}{2}}$$

waarbij $||\cdot||_2$ de Euclidische norm op $\mathbb{R}^n$ is,

$$||v||_2 = \left(\sum_{i=1}^n v_i^2 \right)^{\frac{1}{2}}, \quad v = (v_1, \dots, v_n) \in \mathbb{R}^n,$$

gebruiken, omdat deze van een inproduct afgeleide norm makkelijk hanteerbaar is voor minimaliseringsproblemen. Op de belangrijke vraag naar de relatie tussen de zo gevonden schatting par_x en de parametervector par_x zullen we hier niet verder ingaan; we verwijzen hiervoor naar de literatuur, bijv. Heineken e.a. [1967].

In deze sectie hebben wij het niet ondubbelzinnig geformuleerde probleem (5.6) om een schatting van par_x te maken in verband gebracht met het veel concretere probleem:

Vindt de waarde(n) par_s van de parametervector zo, dat

$$\begin{aligned}
 \text{Res}(\text{par}_s) &= \left(\sum_{t \in T} \| \text{obs}(t) - x(t, \text{par}_s) \|^2 \right)^{\frac{1}{2}} \\
 (5.8) \qquad &= \inf_{\text{par}} \text{Res}(\text{par}) \\
 &= \inf_{\text{par}} \left(\sum_{t \in T} \| \text{obs}(t) - x(t, \text{par}) \|^2 \right)^{\frac{1}{2}}
 \end{aligned}$$

dat wil zeggen: vindt de parameters par_s , die $\text{Res}(\text{par})$ minimaliseren.

Minimalisering van $\text{Res}(\text{par})$

Voor het geval dat $x(t, \text{par})$ een lineaire functie van par is, dat wil zeggen $x(t, \text{par}) = x_0(t) + x_1(t)\text{par}$, waarbij x_0 respectievelijk x_1 een van de tijd t afhankelijke vector uit $\mathbb{R}^n$ respectievelijk lineaire afbeelding $\mathbb{R}^m \rightarrow \mathbb{R}^n$ zijn, is het residu $\text{Res}(\text{par})$ een quadratische vorm in par . Het minimaliseren van $\text{Res}(\text{par})$ is dan eenvoudig uit te voeren met de gebruikelijke methoden uit de numerieke algebra.

Wanneer $\text{Res}(\text{par})$ geen quadratische vorm in par is, zijn we voor de oplossing van probleem (5.8) aangewezen op een iteratieve methode. Bij de keuze van zo'n iteratieve procedure moet in de overwegingen betrokken worden, dat elke functie-evaluatie van $\text{Res}(\text{par})$ tamelijk duur is, aangezien deze het numeriek integreren van het beginwaardeprobleem (5.1) met zich meebrengt. Dit impliceert, dat bij voorkeur van een snel convergent proces gebruik gemaakt moet worden. Bovendien volgt uit bovenstaande opmerking, dat een belangrijk deel van de aandacht gericht moet worden op het vinden van een efficiënte aan de aard van het probleem aangepaste numerieke integratiemethode voor de berekening van $x(t, \text{par})$.

Wij hebben het volgende proces gekozen:

Bij de n -de stap, wanneer par_n bekend is, wordt $x(t, \text{par})$ in par_n gelineariseerd; dat wil zeggen, dat we in plaats van $x(t, \text{par})$

$$\bar{x}(t, \text{par}) = x(t, \text{par}_n) + \frac{\partial}{\partial \text{par}} x(t, \text{par}_n)(\text{par} - \text{par}_n),$$

in $\text{Res}(\text{par})$ substitueren. Aangezien $\bar{x}(t, \text{par})$ een lineaire functie van par is, kunnen we $\bar{\text{Res}}(\text{par})$,

$$\bar{\text{Res}}(\text{par}) = \left(\sum_{t \in T} \|\text{obs}(t) - \bar{x}(t, \text{par})\|_2^2 \right)^{\frac{1}{2}}$$

eenvoudig minimaliseren. Voor par_{n+1} nemen we dan de parametervector z_0 , dat

$$\bar{\text{Res}}(\text{par}_{n+1}) = \inf_{\text{par}} \bar{\text{Res}}(\text{par}).$$

Dit iteratie schema is in wezen niets anders dan een gemodificeerd Newton schema voor de bepaling van het nulpunt van $\frac{d}{d \text{ par}} \bar{\text{Res}}^2(\text{par})$.

Er geldt namelijk:

$$\begin{aligned} \text{a. } \frac{d}{d \text{ par}} \bar{\text{Res}}^2(\text{par})(.) &= -2 \sum_{t \in T} (\text{obs}(t) - x(t, \text{par}), \frac{\partial}{\partial \text{par}} x(t, \text{par})(.)) \\ (5.9) \quad \text{b. } \frac{d^2}{d \text{ par}^2} \bar{\text{Res}}^2(\text{par})(.,.) &= 2 \sum_{t \in T} \left(\frac{\partial}{\partial \text{par}} x(t, \text{par})(.) , \frac{\partial}{\partial \text{par}} x(t, \text{par})(.) \right) \\ &\quad + \sum_{t \in T} (\text{obs}(t) - x(t, \text{par}), - \frac{\partial^2}{\partial \text{par}^2} x(t, \text{par})(.,.)). \end{aligned}$$

Voor de overzichtelijkheid voeren wij de volgende notaties in:

a. de functie

$$\text{par} \mapsto \frac{d}{d \text{ par}} \bar{\text{Res}}^2(\text{par})(.)$$

noemen wij F ; F beeldt $\mathbb{R}^m$ af in de ruimte van lineaire afbeeldingen van $\mathbb{R}^m$ in $\mathbb{R}$.

b. de functie

$$\text{par} \rightarrow \frac{d^2}{d \text{par}^2} \text{Res}^2(\text{par})(\dots)$$

noemen wij DF, aangezien het de afgeleide is van F uit a.

c. de functie

$$\text{par} \rightarrow 2 \sum_{t \in T} \left(\frac{\partial}{\partial \text{par}} x(t, \text{par})(\dots) , \frac{\partial}{\partial \text{par}} x(t, \text{par})(\dots) \right)$$

noemen wij $\overline{DF}$; $\overline{DF}$ beeldt $\mathbb{R}^m$ af in de ruimte van bilineaire afbeeldingen van $\mathbb{R}^m$ in $\mathbb{R}$.

d. de functie

$$\text{par} \rightarrow \sum_{t \in T} (\text{obs}(t) - x(t, \text{par})) , -2 \frac{\partial^2}{\partial \text{par}^2} x(t, \text{par})(\dots)$$

noemen wij G; G beeldt $\mathbb{R}^m$ af in de ruimte van de bilineaire afbeeldingen van $\mathbb{R}^m$ in $\mathbb{R}$.

Het Newton iteratie schema voor de bepaling van het nulpunt van F wordt beschreven door de formule

$$(5.10) \quad \text{par}_{n+1} = \text{par}_n - (DF(\text{par}_n))^{-1} F(\text{par}_n)^*.$$

In het hierboven door ons voorgestelde iteratieproces wordt par_{n+1} uit par_n bepaald volgens de formule

$$(5.11) \quad \text{par}_{n+1} = \text{par}_n - (\overline{DF}(\text{par}_n))^{-1} F(\text{par}_n)^*.$$

* Met $(DF(\text{par}_n))^{-1}$ is bedoeld de inverse van de lineaire operator

$$\text{par} \rightarrow DF(\text{par}_n)(\text{par}, \dots) ; \text{ evenzo voor } (\overline{DF}(\text{par}_n))^{-1}.$$

Dit is onmiddellijk in te zien door

$$\begin{aligned} \overline{\text{Res}}^2(\text{par}) \\ &= \sum_{t \in T} \left\| \text{obs}(t) - x(t, \text{par}_n) - \frac{\partial}{\partial \text{par}} x(t, \text{par}_n)(\text{par} - \text{par}_n) \right\|_2^2 \\ &= \overline{\text{Res}}(\text{par}_n) + F(\text{par}_n)(\text{par} - \text{par}_n) + \overline{\text{DF}}(\text{par}_n)(\text{par} - \text{par}_n, \text{par} - \text{par}_n) \end{aligned}$$

naar par te differentiëren (om zo de minimaliserende $\text{par} = \text{par}_{n+1}$ te vinden). Uit (5.10) en (5.11) volgt onder de vooronderstelling, dat $\text{obs}(t) - x(t, \text{par}_x) = 0$ onmiddellijk, dat (5.11) een gemodificeerd Newton proces is (zie Vainberg [1964], pag. 258 e.v.), waaruit we kunnen concluderen dat dan het proces (5.11) voor startwaarden par_0 in omgeving van par_x convergeert naar par_x en dat de convergentie quadratisch is. Met de gebruikelijke methoden kan voor het algemene geval $\text{obs}(t) - x(t, \text{par}_x) \neq 0$, bewezen worden, dat voor voldoende kleine $\|\text{obs}(t) - x(t, \text{par}_x)\|$ het proces convergeert voor startwaarden par_0 in een omgeving van par_x en dat lineaire convergentie met een convergentiefactor evenredig aan $\|\text{obs}(t) - x(t, \text{par}_x)\|$ optreedt; aangezien $\|\text{obs}(t) - x(t, \text{par}_x)\|$ over het algemeen tamelijk klein is, zal zeer snelle convergentie optreden in een omgeving van par_x .

Berekening van $x(t, \text{par})$ en $\frac{\partial}{\partial \text{par}} x(t, \text{par})$

Bij de in de vorige sectie voorgestelde methode om $\text{Res}(\text{par})$ te minimaliseren, moet niet alleen $x(t, \text{par})$ maar ook $\frac{\partial}{\partial \text{par}} x(t, \text{par})$ bij elke iteratieslag berekend worden. Hiertoe zijn twee voor de hand liggende methoden te noemen:

- a. numerieke differentiatie.
- b. aangezien $\frac{\partial}{\partial \text{par}} x(t, \text{par})$ oplossing is van het beginwaardeprobleem

$$\begin{aligned}
 (5.12) \quad & \frac{d}{dt} \frac{\partial}{\partial \text{par}} x(t, \text{par}) \\
 &= \frac{\partial}{\partial x} f(x(t, \text{par}), t, \text{par}) \frac{\partial}{\partial \text{par}} x(t, \text{par}) + \\
 & \quad \frac{\partial}{\partial \text{par}} f(x(t, \text{par}), t, \text{par}),
 \end{aligned}$$

met beginvoorwaarde

$$\frac{\partial}{\partial \text{par}} x(t_0, \text{par}) = 0$$

kan de berekening plaatsvinden door numerieke integratie van het gecombineerde stelsel (5.1) - (5.12).

Hoewel de onder a genoemde methode veelal aanzienlijk minder rekenwerk vraagt, is toch de onder b genoemde methode gekozen, wegens de problemen, die numerieke differentiatie met zich meebrengt. Voor het numeriek oplossen van het stelsel differentiaalvergelijkingen (5.1) - (5.12) is het van belang de jacobiaan eens nader te bezien.

Deze is:

$$J_c = \begin{pmatrix} \frac{\partial f}{\partial x} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ \frac{\partial}{\partial x} \frac{\partial f}{\partial \text{par}_1} & \frac{\partial f}{\partial x} & & & \vdots \\ \vdots & 0 & \ddots & & 0 \\ \frac{\partial}{\partial x} \frac{\partial f}{\partial \text{par}_m} & 0 & \cdots & \cdots & \frac{\partial f}{\partial x} \end{pmatrix}$$

Hieruit is direct in te zien, dat de eigenwaarden van de jacobiaan J_c dezelfde zijn als de eigenwaarden van de jacobiaan $\frac{\partial f}{\partial x}$ van (5.1). Dit betekent dat het gecombineerde stelsel even stijf is als het oorspronkelijke stelsel (5.1).

Toepassing op ESCEP

Om het voorgaande toe te kunnen passen op het schatten van de parameters in ESCEP moet men de beschikking hebben over een numerieke integratie methode, die ervoor geschikt is het stelsel differentiaalvergelijkingen (5.5) - (5.12) op te lossen. Aangezien we veronderstellen dat de observaties met een fout van ongeveer één procent belast zijn, hoeven we geen grote nauwkeurigheid te eisen. Het is echter wel zaak een efficiënte methode te gebruiken, omdat het stelsel (5.5) - (5.12) tijdens het iteratieve proces om $\text{Res}(\text{par})$ te minimaliseren, herhaaldelijk opgelost moet worden. Gezien bovenstaande opmerkingen en het feit, dat stelsel (5.5) veelal stijf is (zie hoofdstuk 1), bestaat het gevaar, dat door de stabiliteit een veel kleinere stapgrootte wordt voorgeschreven dan nodig is om aan de nauwkeurigheidseisen te voldoen.

Uit hetgeen in hoofdstuk 3 is besproken en het feit, dat beide stelsels dezelfde eigenwaarden hebben, volgt dat de stelsels (5.5) en (5.5) - (5.12) zich, wat de stabiliteit betreft, op dezelfde manier zullen gedragen. Daarom bekijken we eerst stelsel (5.5).

De jacobiaan hiervan is

$$(5.13) \quad J_F = \begin{pmatrix} c-1 & \text{par}_1+s \\ \text{par}_2(1-c) & -\text{par}_2(\text{par}_3+s) \end{pmatrix}$$

Met behulp van de stelling van Gerschgorin vinden we de schatting

$$(5.14) \quad \begin{aligned} \lambda_1 &= c-1 + \delta_1, \quad |\delta_1| < \text{par}_1 + s \\ \lambda_2 &= -\text{par}_2(\text{par}_3+s) + \delta_2, \quad |\delta_2| < \text{par}_1+s, \end{aligned}$$

voor de eigenwaarden λ_1 en λ_2 van J_F . Uit (5.13) is bovendien direct af te leiden dat λ_1 en λ_2 beide negatief zijn.

Wanneer we voor par_1 , par_2 en par_3 de realistische waarden 0,8 , 1000 en 0,9 in (5.14) substitueren, vinden we

$$\lambda_1 = c - 1 + \delta_1, \quad |\delta_1| < 1,8$$

$$\lambda_2 = -1000(0,9 + s) + \delta_2, \quad |\delta_2| < 1,8.$$

Uit het feit dat c en s oplossingen zijn van het beginwaardeprobleem (5.5), leiden we af, dat $0 \leq c < 1$, $0 < s \leq 1$ en $s(c) = 1$. Dus ligt de kleinste eigenwaarde λ_1 in het interval $[-2,8, 0]$ en de grootste, λ_2 , in het interval $[-1900 - 1,8, -1900 + 1,8]$ wanneer t nul is, en in het interval $[-900 - 1,8, -900 + 1,8]$ wanneer $t \rightarrow \infty$ (dat wil zeggen $s \rightarrow 0$). Dit maakt, dat het toepassen van de klassieke vierpunts vierde orde Runge-Kutta formule een kostbare aangelegenheid is. Uit stelling 3.3. vinden we namelijk, dat door stabiliteitsoverwegingen een staplengte

$$(5.15) \quad \tau < \frac{2,8}{\lambda_2}$$

voorgeschreven wordt; hierbij is 2,8 de reële stabiliteitsgrens van het stabiliteitspolynoom

$$P_4(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24}$$

van de Runge Kutta formule (zie tabel 3.6).

We zullen dus moeten uitzien naar een andere integratiemethode, die beter tegen de gegeven ligging van de eigenwaarden van J_f opgewassen is. Hierbij zal het ongetwijfeld voordelig zijn om ervan gebruik te maken, dat de eigenwaarden λ_1 en λ_2 gelegen zijn in twee goed te localiseren relatief kleine circeltjes. Dit alles wijst in de richting van een exponentieel gefitte methode; tenminste bij gebruik van expliciete methoden. Daarom is als integratiemethode gekozen de door het derde graads exponentieel gefitte polynoom $R_3(z)$ (zie (3.66) en van der Houwen [1970] 4.2) gegenereerde Taylormethode. Het stabiliteitsgebied van $R_3(z)$ bestaat uit de volgende twee delen:

- i. Een cirkel G_1 met een nog vrij te kiezen middelpunt z_0 op de negatieve reële as en diameter d . Wanneer geldt $z_0 \ll -1$, dan is d gelijk aan 2.
- ii. Een bij de oorsprong gelegen gebied G_2 , dat voor $z_0 \ll -1$ nagenoeg gelijk is aan het stabiliteitsgebied van de Euler methode.

Volgens stelling 3.3. zullen we nu uit stabiliteits overwegingen de stap-lengte τ zo kiezen, dat geldt

$$(5.16) \quad \tau \lambda_1 \in G_2 \text{ en } \tau \lambda_2 \in G_1$$

Wanneer we de parameters in $R_3(z)$ zo kiezen, dat het middelpunt z van de cirkel G_1 gelijk is aan $-\tau 10^3(0.9+s)$ (zie 5.14), dan is aan (5.16) voldaan, zodra geldt

$$(5.17) \quad \tau < 1/1.8 \text{ (zie fig. 5.1)).}$$

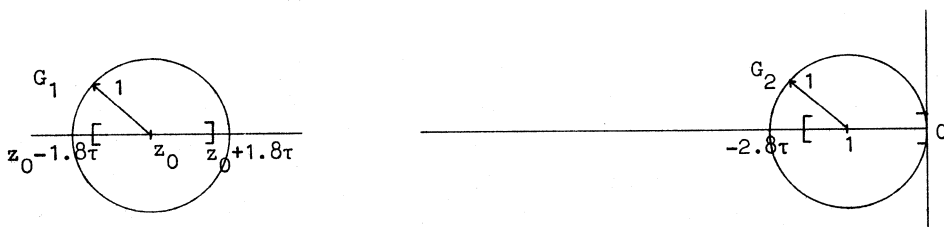


fig. (5.1)

Dit levert een veel gunstiger, door de stabiliteit voorgeschreven, stap-lengte op dan (5.15). Nu zal τ over het algemeen door de geeiste nauwkeurigheid bepaald worden, vooral omdat de genoemde Taylor methode slechts eerste orde consistent is. Bij de numerieke experimenten uit de volgende sectie is gebruik gemaakt van een op het Mathematisch Centrum ontworpen algol-60 procedure van de exponentieel gefitte Taylor methode (zie van der Houwen e.a. [1971]).

Numerieke experimenten

Om het schatten van de parameters in ESCEP met behulp van de geschetste methode te simuleren hebben we op de volgende manier 'observaties' gegenereerd: Eerst is voor een aantal tijdstippen $t \in T$ de oplossing $(s(t), c(t))$ van (5.5), waarbij voor par de te schatten parameterwaarde was gesubstitueerd, met behulp van de in de vorige sectie genoemde integratiemethode met grote nauwkeurigheid berekend. Vervolgens werd elk van deze berekende waarden verstoord door er een homogeen verdeeld random getal uit het interval $[-1/100, 1/100]$ bij op te tellen. De zo verkregen verstoorde waarden van $s(t)$ en $c(t)$, $t \in T$, werden als 'observaties' $\text{obs}(t)$ gebruikt. Op deze manier zijn een aantal experimenten uitgevoerd voor verschillende verzamelingen T en voor verschillende waarden van de te schatten parameters par . Bij deze experimenten werd om rekentijd te sparen $\frac{\partial x}{\partial \text{par}}$ slechts bij elke tweede iteratieslag uitgerekend en bij de daartussen liggende iteratieslagen werd de laatst bekende $\frac{\partial x}{\partial \text{par}}$ gebruikt. Hieronder volgen de resultaten van enige van deze experimenten.

Bij het eerste experiment werden 'observaties' van $s(t)$ en $c(t)$ op de tijdstippen $t \in T$, $T = \{1, 2, 3, \dots, 10\}$, gegeven. De parametervector, die bepaald moest worden, had de componenten $\text{par}_1 = 0.8$, $\text{par}_2 = 1000$, $\text{par}_3 = 0.9$. De stelsels differentiaalvergelijkingen (5.5) en (5.5) - (5.12) werden tijdens het iteratieproces met absolute en relatieve nauwkeurigheid $\eta = 10^{-2}$ opgelost. In de onderstaande tabel is het verloop van het iteratieproces weergegeven door bij de n -de iteratieslag te vermelden:

- a. De berekende parameterwaarde $\text{par}^n = (\text{par}_1^n, \text{par}_2^n, \text{par}_3^n)$,
- b. Het residu $\text{Res}(\text{par}^n)$; waarbij we een genormeerde vorm van het residu gebruikt hebben; namelijk

$$\text{Res}(\text{par}^n) = \left(\frac{1}{|T|} \sum_{t \in T} \|\text{obs}(t) - x(t, \text{par}^n)\|_2^2 \right)^{\frac{1}{2}}$$

- c. Het aantal stappen, k_n , nodig voor de integratie van stelsel differentiaalvergelijkingen bij de opgegeven nauwkeurigheid η .

Tabel (5.1)

n	par_1^n	par_2^n	par_3^n	$\text{Res}(\text{par}^n)$	k_n
0	0,400	10	0,400	0,34	18
1	0,731	16	0,799	0,0703	17
2	0,782	20	0,868	0,0242	26
3	0,800	44	0,900	0,0152	24
4	0,800	58	0,900	0,0123	38
5	0,800	157	0,900	0,0103	24
6	0,800	219	0,901	0,00886	53
7	0,800	1024	0,901	0,00880	42
8	0,800	16664	0,901	0,00827	97
9	0,800	37597	0,901	0,00843	120

$T = \{1, 2, 3, \dots, 10\}.$

$\text{par} = \{0.8, 1000, 0.9\}.$

$n = 10^{-2}.$

Bij het tweede experiment is voor T , par en η het volgende gekozen:

$$T = \{0.001, 0.01, 0.1, 1, 2, \dots, 7\},$$

$$\text{par} = (0.8, 1000, 0.9)$$

$$\eta = 10^{-4}.$$

Tabel (5.2)

n	par_1^n	par_2^n	par_3^n	$\text{Res}(\text{par}^n)$	k_n
0	0,400	10	0,400	0,308	23
1	0,727	18	0,802	0,185	38
2	0,784	24	0,878	0,174	76
3	0,827	83	0,938	0,127	61
4	0,817	109	0,923	0,116	126
5	0,846	291	0,948	0,0751	85
6	0,836	390	0,936	0,0585	193
7	0,803	774	0,903	0,0175	117
8	0,802	882	0,903	0,0115	257
9	0,802	1032	0,902	0,00739	127

Dit experiment wordt herhaald met $\eta = 10^{-2}$.

Tabel (5.3)

n	par_1^n	par_2^n	par_3^n	$\text{Res}(\text{par}^n)$	k_n
0	0,400	10	0,400	0,307	31
1	0,727	18	0,802	0,186	13
2	0,785	24	0,878	0,174	19
3	0,829	83	0,939	0,128	24
4	0,818	109	0,924	0,117	36
5	0,846	287	0,948	0,0757	33
6	0,836	386	0,937	0,0591	55
7	0,803	771	0,903	0,0178	34
8	0,803	879	0,903	0,0116	61
9	0,803	1031	0,902	0,0074	28

Literatuur

- | | |
|----------------------|--|
| Heineken, F.G. | On the accuracy of determining rate constants in enzymatic reactions. |
| Tsuchiya, M. | Math. Biosc. 2 (1967) 115. |
| Aris, R. | |
| Houwen, P.J. van der | One step methods for linear initial value problems II, Applications to stiff equations, Report TW 122, Mathematisch Centrum, Amsterdam (1970). |
| Houwen, P.J. van der | One step methods for linear initial value problems III, Numerical results, Report TW 130, Mathematisch Centrum, Amsterdam (1971). |
| Beentjes, P. | |
| Dekker, K. | |
| Slagt, E | |
| Lee, E.S. | Quasilinearisation and invariant imbedding. Academic Press, New York, 1968. |
| Vainberg, M.M. | Variational methods for the study of nonlinear operators. Holden-Day, Inc., San Francisco, 1964. |

6. Fout-formules en stapkeuze-strategieën

Bij het numeriek oplossen van differentiaalvergelijkingen zal men in de regel een idee willen hebben over de nauwkeurigheid van de verkregen oplossing. Als de oplossing ook analytisch te verkrijgen is, maar in een onaantrekkelijke vorm voor directe berekening, dan heeft men toch in de analytische uitdrukking een middel ter controle van de numerieke oplossing.

Beschouw bijvoorbeeld de oplossing $y(x) = e^{-x^2} \int_0^x e^{t^2} dt$ van de vergelijking $y' = 1-2xy$; gesteld dat $y(x)$ op $[a,b]$ in een flink aantal punten berekend moet worden, dan kan het raadzaam zijn de differentiaalvergelijking numeriek op te lossen en de nauwkeurigheid na een aantal stappen te controleren door y met behulp van een quadratuur-formule nauwkeurig te berekenen.

Is echter geen analytische oplossing bekend -waar we in het vervolg van uit zullen gaan- dan moeten we methoden vinden om een min of meer betrouwbare foutschatting te maken. De gemaakte fout hangt in het algemeen sterk af van de grootte van de integratie-stap, zodat men met behulp van een goede stapkeuze-strategie de fout binnen de door de probleemsteller opgegeven grenzen kan houden.

In dit hoofdstuk zullen we ons voornamelijk bezig houden met het ontwerpen van fout-formules en stapkeuze-strategieën van eenstapsformules van de vorm

$$(6.1) \quad y_{n+1} = y_n + h_n \phi(x_n, y_n; h_n).$$

De differentiaalvergelijkingen die we beschouwen zijn van het type

$$(6.2) \quad y' = f(x,y), \quad y(a) = y_0.$$

In het vervolg zullen we ons houden aan de notaties van hoofdstuk 3.

Discrepantie en discretiseringsfout

We onderscheiden eerst een aantal typen van fouten die bij de bereke-

ning van de oplossing van (6.2) met methode (6.1) optreden, te weten *discrepantie*, *discretiseringsfout* en *afrondingsfout*.

Over afrondingsfouten zullen we het in dit hoofdstuk niet hebben; we nemen aan dat hun effect, zo nodig gereduceerd door meervoudige precisie berekeningen, gering is.

We gaan nu nader in op de discretiseringsfout ϵ_n , gedefinieerd door (3.26).

Henrici (1962) geeft voor deze fout het volgende asymptotische gedrag aan

$$(6.3) \quad \epsilon_n = e(x_n) h^p + O(h^{p+1}),$$

waarin de functie e de oplossing is van het beginwaarde-probleem

$$(6.4) \quad \begin{cases} e'(x) = f_y(x, y(x)) e(x) + [\theta(x)]^p \phi(x, y(x)), \\ e(a) = 0 \end{cases}.$$

Hierin wordt ϕ , de *principale foutfunctie*, gedefinieerd door de vergelijking

$$(6.5) \quad \rho(x, y; h) = h^{p+1} \phi(x, y) + O(h^{p+2}),$$

terwijl de functie θ gegeven wordt door

$$\theta(x_n) = h_n/h, \quad 0 < \theta(x) \leq 1, \quad n = 0, 1, 2, \dots,$$

waarin h een constante maximale stapgrootte is.

Ter illustratie geven we nu een toepassing van formule (6.3).

Als $y(x, h)$ de numerieke oplossing (in het punt x) en h de gebruikte stapgrootte is, dan geldt volgens (6.3)

$$y(x, h) = y(x) + h^p e(x) + O(h^{p+1}),$$

en

$$y(x, qh) = y(x) + (qh)^p e(x) + O(h^{p+1}), \quad q \neq 0, 1.$$

Elimineren we $e(x)$ uit de twee bovenstaande formules dan krijgen we

$$y(x) = \frac{y(x, qh) - q^p y(x, h)}{1 - q^p} + O(h^{p+1}).$$

Deze zogenaamde *Richardson extrapolatie naar $h = 0$* benadert dus de exacte oplossing met een fout waarvan de orde één hoger is dan de orde van de integratie-methode.

In het vervolg blijft dit Richardson proces echter buiten beschouwing omdat we ons tot doel stellen om het interval waarop de oplossing gewenst wordt slechts eenmaal te integreren. Tevens zullen we niet aan zogenaamde stapverwerping doen.

Een alternatief voor de berekening van ϵ_n zou in het oplossen van (6.4) kunnen liggen, waarin we (bij gebrek aan beter) voor $y(x_n)$ de waarde van y_n zouden kunnen nemen. Dit vervalt echter omdat we geen uitdrukking voor ϕ hebben.

Om althans een schatting van de locale nauwkeurigheid te maken, beperken we ons tot een berekening van de discrepantie ρ_n .

Voor een aantal eenstapsmethoden te weten, klassieke Runge-Kutta methoden, gestabiliseerde Runge-Kutta methoden, Taylor-reeks-methoden en methoden voor stijve differentiaalvergelijkingen zullen we ons nu gaan bezighouden met de berekening van de discrepantie ρ_n .

In het vervolg zullen we schattingen van de discrepantie met ρ_n^* aangeven.

Klassieke Runge-Kutta methoden

Klassieke Runge-Kutta methoden zijn meestal ontworpen om met een minimaal aantal functie-evaluaties m (zie (3.10)) een zo hoog mogelijke orde van consistentie p te verkrijgen.

Bij een gegeven schema is het in principe mogelijk om, met extra evaluaties, ρ_n te berekenen.

Doen we dit echter, dan hebben we een $(p+1)^e$ orde methode ontworpen n.l. door stapsgewijs als verbeterde oplossing

$$y_{n+1} + \rho_n^*,$$

te nemen, waarmee we ρ_n^* a.h.w. tot een zogenaamde laatste correctie-term (l.c.t.) degraderen.

Zoals genoemd in hoofdstuk 2, gebruikt Zonneveld (1964) voor de berekening van de l.c.t. één extra evaluatie.

Sommige Runge-Kutta formules hebben een zodanige structuur dat er meerdere benaderingen van y_{n+1} mee gemaakt kunnen worden die verschillend van orde zijn. Het verschil tussen de twee hoogste orde benaderingen van y_{n+1} geeft dan een fout die, wat orde betreft, vergelijkbaar is met de l.c.t.. Als voorbeeld geven we de volgende formule van Sarafyan (zie Lapidus & Seinfeld (1971)).

$$(6.6) \quad \begin{pmatrix} \Lambda \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & & & & & \\ & 1/2 & & & & \\ & 1/4 & 1/4 & & & \\ & 0 & -1 & 2 & & \\ & 7/27 & 10/27 & 0 & 1/27 & \\ & 28/625 & -125/625 & 546/625 & 54/625 & -378/625 \\ & 14/336 & 0 & 0 & 35/336 & 162/336 & 125/336 \end{pmatrix}$$

Zoals bekend kunnen met deze matrix de vectoren $k_n^{(0)}, \dots, k_n^{(5)}$ berekend worden (zie formule 3.10)).

De waarde van y_{n+1} kan met behulp van (6.6) op verschillende manieren worden berekend:

$$\begin{cases} y_{n+1}^1 = y_n + k_n^{(0)}, p = 1, \\ y_{n+1}^2 = y_n + k_n^{(1)}, p = 2, \\ y_{n+1}^4 = y_n + (k_n^{(0)} + 4k_n^{(2)} + k_n^{(3)})/6, p = 4, \\ y_{n+1}^5 = y_n + (14k_n^{(0)} + 35k_n^{(3)} + 162k_n^{(4)} + 125k_n^{(5)})/336, p = 5. \end{cases}$$

De laatste formule volgt per definitie uit (3.10).

Met behulp van y_{n+1}^4 en y_{n+1}^5 kunnen we de volgende fout-schatting maken

$$(6.7) \quad \rho_n^* = y_{n+1}^5 - y_{n+1}^4 = O(h^5).$$

In tabel 6.1 geven we enige relevante getallen die we verkregen door de vergelijking

$$y' = -y^2, y(0) = 1,$$

met de 5^e orde Runge-Kutta methoden van Zonneveld resp. Sarafyan op te lossen.

Voor ρ_n^* namen we bij Zonneveld de l.c.t. en bij Sarafyan de fout gegeven door (6.7).

De discrepantie ρ_n is voor deze eenvoudige vergelijking direct te berekenen.

Tabel 6.1 Resultaten van de formules van Zonneveld en Sarafyan toegepast op $y' = -y^2, y(0) = 1$

stap	Zonneveld			Sarafyan		
	$ \rho_n^*/\rho_n _{\max.}$	$ \rho_n^*/\rho_n _{\min.}$	absolute fout op $x = 5$	$ \rho_n^*/\rho_n _{\max.}$	$ \rho_n^*/\rho_n _{\min.}$	absolute fout op $x = 5$
.5	1200	25	2.0×10^{-5}	7300	20	3.4×10^{-5}
.1	3000	1000	7.0×10^{-10}	5×10^6	5000	7.8×10^{-9}
.05	6000	2500	1.7×10^{-10}	3×10^6	8×10^4	2.3×10^{-10}

Gestabiliseerde Runge-Kutta methoden

Voor deze formules, waarbij de nadruk op stabiliteitseigenschappen ligt, zijn in het algemeen meer functie-evaluaties per stap nodig dan bij ongestabiliseerde Runge-Kutta formules van dezelfde orde. Dit grotere aantal functie-evaluaties maakt echter in vele gevallen een schatting van een l.c.t., of soms zelfs van de discrepantie, direct mogelijk.

Als voorbeeld beschouwen we de 3-punts gestabiliseerde Heun-formule (3.14).

De l.c.t. kan onder andere door een drietal formules worden gegeven

$$\rho_n^* = \begin{cases} k_2 - k_0, \\ 4k_1 - 4k_0, \\ \frac{4}{3}(k_2 - k_1). \end{cases}$$

Voor lineaire vergelijkingen kunnen we nog de volgende fout-schatting maken

$$\rho_n^* = (15k_0 + 5k_2 - 20k_1)/3 = O(h^3).$$

Uit de bovenstaande formules blijkt dat ρ_n^* niet door een eenduidige formule bepaald wordt. De oorzaak ligt in het feit dat de gewichten

θ_j' ($j = 0, \dots, m-1$) van de fout-formule

$$(6.8) \quad \rho_n^* = \sum_{j=0}^{m-1} \theta_j' k_n^{(j)}$$

de onbekenden zijn van een onderbepaald stelsel lineaire vergelijkingen (zie van der Houwen (1971a)).

In het algemeen variëren de mogelijke oplossingen van dat stelsel nogal sterk in norm. Aangezien een van de vergelijkingen de vorm

$$\sum_{j=0}^{m-1} \theta_j' = 0 \text{ heeft, ligt het voor de hand om die oplossing te kiezen die}$$

blijkens formule (6.8) zo weinig mogelijk cijferverlies zou kunnen geven.

Dit kunnen we bereiken door

$$\sum_{j=0}^{m-1} |\theta'_j|$$

te minimaliseren. Het vinden van de gewichten θ'_j wordt hiermee teruggebracht tot het oplossen van een lineair-programmeringsprobleem.

Als voorbeeld geven we de gewichten θ'_j ($j = 0, \dots, 6$) voor een 7-punts Runge-Kutta formule van de orde 3. We vinden de volgende oplossing met behulp van lineaire programmering:

$$\theta'_0 = \theta'_3 = \theta'_2 = 0, \theta'_1 = 1.79, \theta'_4 = -4.01, \theta'_5 = 1.69, \theta'_6 = .53 .$$

Stellen we daarentegen $\theta'_4 = \theta'_5 = \theta'_6 = 0$ dan krijgen we:

$$\theta'_0 = -2704, \theta'_1 = 2439, \theta'_2 = 278, \theta'_3 = -13 .$$

In tabel 6.2 geven we enige resultaten die we verkregen door deze 7-punts Runge-Kutta formule, met de genoemde twee mogelijkheden van foutgewichten, toe te passen op de volgende vergelijking

$$\begin{cases} y' = -200(y-F(x)) + F'(x) , \\ y(0) = 10 , \\ F(x) = 10 - (10+x)e^{-x} . \end{cases}$$

Onder II geven we in de tabel een schatting ρ_n^* van de discrepantie met behulp van de gewichten, verkregen met lineaire programmering; onder I zijn voor ρ_n^* de andere vermelde foutgewichten gebruikt.

Tabel 6.2 Foutschattingen met verschillende foutgewichten
voor een 7-punts Runge-Kutta formule van de orde 3.

stap	absolute fout op $x=10$	I	II
		$ \rho_n^*/\rho_n _{\text{gemidd.}}$	$ \rho_n^*/\rho_n _{\text{gemidd.}}$
0.1	4.9×10^{-4}	2000	80
.05	9.7×10^{-6}	200	32
.025	1.1×10^{-7}	67	26

Taylor-reeks methoden

Voor deze methoden gelden, met betrekking tot de berekenbaarheid van de afbreekfout, dezelfde overwegingen als bij Runge-Kutta methoden.

Voor ongestabiliseerde methoden is de discrepantie

$$\rho_n = - \frac{1}{(p+1)!} f^{(p)}(x,y) h_n^{p+1} + O(h_n^{p+2}).$$

Berekening van ρ_n kan deze in principe weer tot een laatste correctieterm degraderen.

Voor gestabiliseerde Taylor methoden kunnen we in het algemeen wel een schatting van ρ_n maken; deze is dan

$$\rho_n^* = \left(\frac{1}{(p+1)!} - \beta_{p+1} \right) f^{(p)}(x,y) h_n^{p+1},$$

terwijl een betere schatting mogelijk is als m_2 (zie (3.5)) groter is dan $p+1$.

Methoden voor stijve differentiaalvergelijkingen

Deze methoden, die o.a. in hoofdstuk 3 zijn behandeld, vereisen een geheel andere aanpak voor een schatting van de fout zoals tot nu toe gedaan is.

Bij stijve vergelijkingen worden de hogere afgeleiden van de oplossing in hoofdzaak bepaald door snel variërende functies die corresponderen met de in modulus grote eigenwaarden van de Jacobiaan, waardoor een fout-berekening op de manier zoals we tot nu toe hebben voorgesteld, praktisch geen informatie geeft over de werkelijke waarde van de discrepantie.

We geven daarom een geheel andere techniek om fouten te schatten, die beschreven staat in van der Houwen (1971b).

In plaats van de lokaal analytische oplossing in het differentie-schema te substitueren en daarna eindig veel termen van Taylor-reeksen te vergelijken, bekijken we nu hoe goed de numerieke oplossing op (x_n, x_{n+1}) , die we zullen aangeven met $y_n(x)$, aan de differentiaalvergelijking voldoet,

met andere woorden we beschouwen het *residu*:

$$(6.9) \quad \zeta_n(\xi) = \frac{\partial}{\partial \xi} y_n(x_n + \xi) - f(x_n + \xi, y_n(x_n + \xi)),$$

waarvoor de volgende stelling geldt.

Stelling 6.1

De discrepantie ρ_n voldoet aan

$$(6.10) \quad \rho_n(h) \sim \int_0^h \zeta_n(\xi) d\xi,$$

vooropgesteld dat

$$I_n(h) = \int_0^h f(x_n + \xi, y(x_n + \xi)) - f(x_n + \xi, y_n(x_n + \xi)) d\xi,$$

te verwaarlozen is ten opzichte van het rechterlid van (5.7).

Verder geldt dan

$$I_n(h) = O(h^{p+2}) \text{ voor } h \rightarrow 0.$$

Bewijs:

$$\begin{aligned} \rho_n(h) &= -y(x_n+h) + y_n(x_n+h) = -y(x_n+h) + y(x_n) - y(x_n) + y_n(x_n+h) \\ &= \int_0^h -\frac{\partial}{\partial \xi} y(x_n + \xi) d\xi + \int_0^h \frac{\partial}{\partial \xi} y_n(x_n + \xi) d\xi \\ &= \int_0^h -f(x_n + \xi, y(x_n + \xi)) + f(x_n + \xi, y_n(x_n + \xi)) d\xi + \\ &\quad + \int_0^h -f(x_n + \xi, y_n(x_n + \xi)) + \frac{\partial}{\partial \xi} y_n(x_n + \xi) d\xi \\ &= \int_0^h \zeta_n(\xi) d\xi - I_n(h). \end{aligned}$$

Omdat

$$y(x_n + \xi) - y_n(x_n + \xi) = O(\xi^{p+1}),$$

volgt uit de Taylor ontwikkeling van $f(x_n + \xi, y(x_n + \xi)) - f(x_n + \xi, y_n(x_n + \xi))$

$$I_n(h) = O(h^{p+2}) \text{ voor } h \rightarrow 0.$$

Met behulp van stelling 6.1 kunnen we de volgende bovengrens geven voor $\rho_n(h)$:

$$(6.11) \quad |\rho_n(h)| \leq h |\zeta_n(h)|.$$

Voor Taylor-methoden kunnen we ζ_n zonder veel moeite berekenen en dus met (6.11) een bovengrens voor ρ_n aangeven.

Voor Runge-Kutta methoden is de berekening van ζ_n praktisch onmogelijk. We kunnen echter met behulp van de trapezium-regel de volgende (ruwe) schatting van ρ_n uit (6.10) maken:

$$(6.12) \quad \rho_n^* = y_{n+1} - y_n - \frac{1}{2}h [f(x_{n+1}, y_{n+1}) - f(x_n, y_n)].$$

De tot nu toe aangegeven formules voor de berekening van de discrepantie ofwel laatste-correctie-term geven buiten een indicatie omtrent nauwkeurigheid tevens op ieder tijdstip een maat voor de te nemen volgende integratie-stap. We gaan ons nu bezighouden met formules waarmee we de stapgrootte kunnen regelen.

Stapkeuze-strategieën

We gaan er van uit dat we de discrepantie in iedere stap onder controle willen houden, m.a.w. voor iedere integratiestap kiezen we een h_n en een tolerantie η_n zodanig dat

$$(6.13) \quad |\rho_n(h_n)| = \eta_n.$$

De tolerantie η_n kunnen we bijvoorbeeld als volgt berekenen

$$(6.14) \quad \eta_n = \eta_{\text{als}} + \eta_{\text{rel}} |u_k|,$$

waarin we η_{als} en η_{rel} nog, zo nodig, laten variëren.

In formule (6.13) is $\rho_n(h_n)$ echter pas bekend na de n^{e} integratiestap, zodat we op de een of andere manier h_n zullen moeten voorspellen met behulp van discrepanties uit vorige stappen.

De meest triviale stapkeuze-strategie krijgen we als we de principale foutfunctie ϕ , gedefinieerd in (6.5), constant veronderstellen, m.a.w.

$$(6.15) \quad |\rho| = Ch^{p+1}.$$

Voor iedere $n > 1$ geldt dan:

$$h_n = h_{n-1} \sqrt[p+1]{\eta_n / |\rho_{n-1}|}.$$

Als we de principale foutfunctie lineair resp. kwadratisch in x veronderstellen dan krijgen we de volgende representaties voor ρ

$$(6.16) \quad |\rho| = (Bx+C)h^{p+1},$$

$$(6.17) \quad |\rho| = (Ax^2+Bx+C)h^{p+1}.$$

Ook uit deze twee formules kan de stapgrootte h_n expliciet worden berekend met behulp van η_n , ρ_{n-1} , ρ_{n-2} , ρ_{n-3} , x_n , x_{n-1} , en x_{n-2} .

Bij stijve vergelijkingen kan men voor een berekening van de stapgrootte bijvoorbeeld uitgaan van de volgende formule

$$(6.18) \quad |\rho| = (Ah+Bx+C)h^{p+1},$$

omdat de principale foutfunctie hier dikwijls afhankelijk blijkt te zijn van de gebruikte stapgrootte.

Met een nulpunts-formule kunnen we h uit (6.18) berekenen. Uit veiligheidsoverwegingen kan men de verkregen stapgrootte nog bijsturen met parameters α en γ op de volgende manier

$$\gamma h_{n-1} \leq h_n \leq \alpha h_{n-1} \quad (\text{bijvoorbeeld: } \alpha=3, \gamma=1.5).$$

In de voorafgaande beschouwingen over stapkeuze-strategieën was de locale nauwkeurigheid ons voornaamste criterium voor de bepaling van de stapgrootte.

We gaan ons nu bezighouden met een methode die erop gericht is om de discretiseringsfout, voor een gegeven aantal stappen, te minimaliseren. (Voor een uitgebreide behandeling van dit probleem: zie Morrison (1962)).

Het uitgangspunt is de (formele) oplossing van (6.4):

$$(6.19) \quad e(x) = u(x) \int_a^x u(t)^{-1} \phi(t, y(t)) [\theta(t)]^k dt,$$

waarin u gegeven wordt door

$$(6.20) \quad u(x) = \exp \int_a^x f_y(t, y(t)) dt.$$

Ons doel is nu om bij een integratie-methode van de orde p , voor een gegeven aantal integratie-stappen n , een stapverdeling $\theta(x)$ op $[a, b]$ te vinden waarbij $e(b)$ minimaal is.

De oplossing van dit probleem wordt onder bepaalde veronderstellingen gegeven door

$$\theta(x) = \frac{A}{g(x)},$$

waarin

$$g(x) = |u(b)u(x)^{-1} \phi(x, y(x))|^{1/(p+1)}$$

en

$$A = \int_a^b g(t) dt.$$

Toepassing:

We beschouwen het probleem

$$y' = y, y(a) = y_0,$$

dat we willen oplossen op $[a,b]$ met een p^e orde integratie-methode. Voor dit speciale geval volgt onmiddellijk dat:

$$u(x) = e^{x-a},$$

en

$$\phi(x,y) = - \frac{y}{(p+1)!}.$$

Hiermee wordt

$$g(x) = \left| - \frac{e^{b-a} y_0}{(p+1)!} \right|^{1/p+1},$$

zodat

$$\theta(x) = b-a.$$

Het gebruik van een constante stapgrootte geeft in dit geval dus de kleinste discretiseringsfout.

Bij een willekeurige differentiaalvergelijking en integratie-methode kunnen we echter voor $\phi(x,y)$ in het algemeen geen analytische uitdrukking vinden, wat de waarde van deze methode aanzienlijk reduceert.

Het is echter eenvoudig aan te tonen dat de methode voor de discrepantie de volgende benadering geeft

$$(6.21) \quad \rho(x,y) \cong \frac{(hA)^{p+1}}{u(b)} u(x),$$

hetgeen een eenvoudige stapkeuze strategie mogelijk maakt.

We kiezen hiertoe een constante c en integreren de vergelijking

$$u' = f_y u, u(a) = 1,$$

met de gegeven differentiaalvergelijking voor y mee.

Als onze benadering van de discrepantie nu groter blijkt te zijn dan $cu(x_n)$ verminderen we de stapgrootte; is de discrepantie kleiner dan $cu(x_n)$ dan maken we de staplengte groter.

Literatuur

- Morrison, D. Optimal mesh size in the numerical integration of an ordinary differential equation.
J. Assoc. Comput. Mach. 9, 98-103 (1962).
- Henrici, P. Discrete variable methods in ordinary differential equations, (1962) Wiley, New York.
- Houwen, P.J. van der Stabilized Runge-Kutta methods with limited storage requirements. Report TW 124/71,
Mathematisch Centrum, Amsterdam (1971a).
- Houwen, P.J. van der One-step methods for linear initial value problems
Beentjes, P. III, Numerical examples. Report TW 130/71
Dekker, K. Mathematisch Centrum, Amsterdam (1971b).
Slagt, E.
- Lapidus & Seinfeld Numerical solution of ordinary differential equations, (1971).
Academic Press, New York.
- Zonneveld, J.A. Automatic numerical integration (1964),
Math. Centre Tracts 8 (Mathematisch Centrum,
Amsterdam).

7. Exponentieel aangepaste methoden

Zoals reeds in hoofdstuk 2 is opgemerkt heeft de Jacobiaan van een stelsel stijve differentiaalvergelijkingen een of meer eigenwaarden met reëel deel veel kleiner dan 0. Hierdoor wordt bij vele integratie-methoden een stringente beperking aan de grootte van de integratie-stap opgelegd. Immers zij $R(z)$ de stabiliteits-functie van de methode, dan impliceert de stabiliteits voorwaarde

$$(7.1) \quad |R(h\delta)| < 1,$$

bij vele expliciete methoden een voorwaarde van de vorm

$$(7.2) \quad |h\delta| < c,$$

waarbij c een van de desbetreffende methode afhankelijke functie van $\arg(\delta)$ is. Aan deze voorwaarden moet dan voor alle eigenwaarden δ van de Jacobiaan voldaan zijn. Het bijbehorende stabiliteits gebied is dan samenhangend, zie bijvoorbeeld de figuren 3.3. en 3.5.

Is echter informatie bekend over de ligging van de eigenwaarden van de Jacobiaan, dan kan $R(z)$ zo gekozen worden dat niet samenhangende stabiliteits gebieden ontstaan; één dichtbij de oorsprong, de andere ver in het linker-halfvlak, zoals al uiteengezet is in de hoofdstukken 2 en 3. Bovendien kunnen we verlangen dat de bij de eigenwaarden met sterk negatief reëel deel behorende component van de oplossing nauwkeurig gerepresenteerd wordt. Dit wordt bereikt door de stabiliteits-functie in de desbetreffende eigenwaarden exponentieel aan te passen. In dit hoofdstuk worden de beschouwingen uit hoofdstuk 2 en 3 uitgebreid.

De methoden van Liniger en Willoughby

Liniger en Willoughby hebben drie impliciete formules ontwikkeld, die de mogelijkheid van exponentiële aanpassing bezitten.

Deze formules, die in hoofdstuk 2 al vermeld werden, worden voor autonome systemen gegeven door

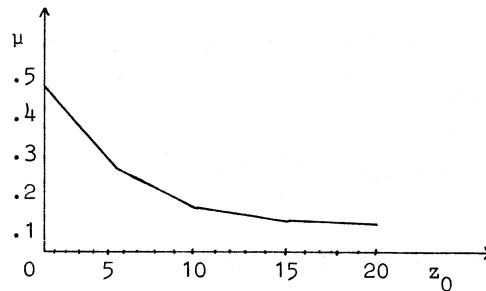
$$(7.3) \quad y_{n+1} = y_n + h[\mu f_n + (1-\mu) f_{n+1}],$$

$$(7.4) \quad y_{n+1} = y_n + \frac{h}{2} (1-a)f_n + (1+a)f_{n+1} \\ + \frac{h^2}{4} [(b-a) J_n f_n - (b+a) J_{n+1} f_{n+1}],$$

$$(7.5) \quad y_{n+1} = y_n + \frac{h}{2} [(1-a)f_n + (1+a)f_{n+1}] \\ + \frac{h^2}{4} [(\frac{1}{3} - a) J_n f_n - (\frac{1}{3} + a) J_{n+1} f_{n+1}],$$

waarbij $J_n = J(x_n, y_n)$.

De formules (7.3), (7.4) en (7.5) worden door Liniger en Willoughby respectievelijk $F^{(1)}$, $F^{(2)}$ en $F^{(3)}$ genoemd. Formule $F^{(1)}$ heeft een orde van nauwkeurigheid 1, van de formules $F^{(2)}$ en $F^{(3)}$ is de orde van nauwkeurigheid 2, respectievelijk 3. De stabiliteitsfuncties $R(z)$ van deze formules zijn in hoofdstuk 3 gegeven (3.69, 3.70, 3.71). In de $F^{(1)}$ formule kan μ (en in de $F^{(3)}$ formule a) zo gekozen worden dat exponentiële aanpassing in een willekeurig negatief reëel punt z_0 bereikt wordt. In figuur 7.1 geven we de grafiek van μ als functie van z_0 . In de $F^{(2)}$ formule kunnen de parameters a en b gekozen worden zodat $R(z_0) = \exp(z_0)$ en $R(z_1) = \exp(z_1)$ voor twee reële of complex toegevoegde punten z_0 en z_1 met negatief reëel deel.

Figuur 7.1

Omdat de beschouwde formules impliciet zijn, leidt toepassing van de formules op een stelsel differentiaalvergelijkingen tot een stelsel vergelijkingen in de componenten van y_{n+1} . Liniger en Willoughby stellen voor om dit stelsel, indien het niet lineair is, op te lossen met de iteratie methode van Newton.

Zou namelijk getracht worden de vergelijkingen door directe iteratie op te lossen, dan komen de moeilijkheden van de kleine tijd constante door de achterdeur weer naar binnen.

Immers de directe iteratie methode levert toegepast op (7.3)

$$y_{n+1}^{(1+1)} = y_n + \mu h f_n + (1-\mu) h f_{n+1}^{(1)};$$

convergentie wordt dan slechts bereikt als $|(1-\mu)h\delta| < 1$ voor alle eigenwaarden δ van de Jacobiaan, en dit is een even sterke beperking van h als bij voorwaarde (7.2).

De iteratie methode van Newton, toegepast op formule $F^{(1)}$ leidt tot het schema

$$(7.6) \quad \left\{ \begin{array}{l} [I - (1-\mu)hJ_{n+1}^{(1)}] \Delta y_{n+1}^{(1)} = y_n - y_{n+1}^{(1)} + \mu h f_n + (1-\mu) h f_{n+1}^{(1)}, \\ \Delta y_{n+1}^{(1)} = y_{n+1}^{(1+1)} - y_{n+1}^{(1)}, \\ y_{n+1}^{(0)} = y_n. \end{array} \right.$$

Voor formule $F^{(2)}$, en ook voor formule $F^{(3)}$ als speciaal geval van formule $F^{(2)}$, gebruiken Liniger en Willoughby een gewijzigde Newton methode:

de term $\frac{dJ}{dy}$ die bij toepassing van de Newton methode zou ontstaan, wordt weggelaten. Dit leidt tot schema

$$\begin{aligned}
 (7.7) \quad & \{4I - 2h(1+a)J_{n+1}^{(1)} + h^2(b+a)[J_{n+1}^{(1)}]^2\} \Delta y_{n+1}^{(1)} = \\
 & = h[2(1+a)I - h(b+a)J_{n+1}^{(1)}]f_{n+1}^{(1)} \\
 & + 4y_n + 2h(1-a)f_n + h^2(b-a)J_n f_n - 4y_{n+1}^{(1)}, \\
 & \Delta y_{n+1}^{(1)} = y_{n+1}^{(1+1)} - y_{n+1}^{(1)}, \\
 & y_{n+1}^{(0)} = y_n.
 \end{aligned}$$

Indien het stelsel differentiaalvergelijkingen lineair is, dan geeft (7.7) in één iteratie de exacte oplossing van de formule $F^{(2)}$.

Voor lineaire inhomogene systemen van het type

$$(7.8) \quad \dot{y} = -Ay + g(x),$$

hebben Liniger en Willoughby het volgende bewezen:

Zij $g(x)$ een polynoom x van de graad p , waarbij p kleiner is of gelijk aan de orde van de beschouwde formule, dan levert toepassing van één der formules $F^{(1)}$, $F^{(2)}$ of $F^{(3)}$ tot een afbreek-fout, die gelijk is aan de afbreek-fout bij het oplossen van het geassocieerde homogene systeem. Dit betekent dat we voor een enkele vergelijking van het type (7.8), waarbij $g(x)$ aan bovengenoemde voorwaarde voldoet, de analytische oplossing in de roosterpunten y_n verkrijgen, indien tenminste in $z_0 = -A.h$ exponentieel gefit is. Dit is opvallend, want de meeste methoden met orde van nauwkeurigheid p geven slechts een exacte oplossing van $\dot{y} = \tilde{g}(x)$, als $\tilde{g}(x)$ een polynoom is van de graad $p-1$, niet van de graad p .

Het is van belang zich af te vragen hoe veel de beschouwde methoden kosten uitgedrukt in aantallen functie- en Jacobiaan-evaluaties. In het algemeen hangt dit af van de strategie die gevolgd wordt bij de iteratieschema's (7.6) of (7.7). Beschouwen we eerst de strategie, waarbij het schema per iteratie stap "exact" behandeld wordt. ("exact" betekent in dit geval dat geen benaderingen in de functie- en Jacobiaan-waarden toegepast worden). De aantallen functie- en Jacobiaan-evaluaties zijn in dit geval gegeven in tabel 7.1.

Tabel 7.1

Aantal functie- en Jacobiaan-evaluaties bij (7.6) en (7.7)

	functie-ev.	Jacobiaan-ev.	
1e iteratie	2	1(7.6)	2(7.7)
volgende iteraties	1	1	

Indien het systeem differentiaalvergelijkingen autonoom is, dan is per iteratieslag slechts één functie-evaluatie en één Jacobiaan-evaluatie vereist. In dat geval geldt namelijk $f(x_{n+1}, y_n) = f(x_n, y_n)$ en $J(x_{n+1}, y_n) = J(x_n, y_n)$. Dat dit bij een niet autonoom systeem niet mogelijk is, zullen we met het volgende voorbeeld illustreren.

Voorbeeld

Beschouw de lineaire vergelijking

$$(7.9) \quad \begin{aligned} \dot{y} &= -y + x, \\ y(0) &= y_0 = 0. \end{aligned}$$

Daar de inhomogene term een polynoom van de graad ≤ 1 is, moet schema (7.6) met één iteratie de exacte oplossing $y(x) = e^{-x} + x - 1$ geven.

Stellen we echter $f_{n+1}^{(0)} = f_n^{(0)}$, dan vinden we voor alle waarden van $h > 0$ $y_1 = 0$.

Schrijven we echter de vergelijking (7.9) als een autonoom stelsel

$$(7.9') \quad \begin{aligned} \dot{y}_1 &= -y_1 + y_2, \quad y_1(0) = y_2(0) = 0, \\ \dot{y}_2 &= 1, \end{aligned}$$

dan geeft schema (7.6) met één functie-evaluatie in één iteratie de exacte oplossing.

Het is natuurlijk wel mogelijk per iteratie slag evaluaties uit te sparen door de schema's (7.6) en (7.7) niet exact te behandelen. In dat geval kunnen we $f_{n+1}^{(0)}$ lineair benaderen:

$$f_{n+1}^{(0)} = f_n^{(0)} + h \frac{\partial f_n}{\partial x}; \text{ bovendien kunnen we de Jacobiaan per stap}$$

constant houden, of per serie stappen. Immers indien het aldus gewijzigde schema (7.6) convergeert, dan voldoet de oplossing aan formule $F^{(1)}$.

Voor schema (7.7) geldt dit echter niet; de Jacobiaan komt daar namelijk ook in het rechterlid voor.

Dit betekent dat formule $F^{(3)}$ dan niet meer 3^e orde exact is.

Een ander bezwaar van het periodiek constant houden van de Jacobiaan is dat de convergentie-snelheid kleiner wordt, zodat wellicht meer iteraties nodig zijn om de oplossing in een voldoende precisie te bereiken.

Numerieke Resultaten

Lapidus en Seinfeld hebben verschillende methoden toegepast op een aantal test voorbeelden van stijve differentiaalvergelijkingen. De methoden van Liniger en Willoughby behaalden daarbij zeer goede resultaten; ondanks grote stappen werd toch een redelijke precisie bereikt.

Een bezwaar van formule $F^{(1)}$ is dat hij door de exponentiële aanpassing slechts een orde van nauwkeurigheid 1 heeft. Kiezen we echter $\mu = \frac{1}{2}$, dan ontstaat de trapezium-regel, en de orde van nauwkeurigheid is 2. Daarom lijkt het in gevallen waarbij de stijve component slechts in het inschakel verschijnsel een rol speelt, geschikter om na verloop van tijd op de trapezium-regel over te gaan. We illustreren dit met het volgende voorbeeld.

Voorbeeld

Beschouw het lineaire stelsel differentiaalvergelijkingen

$$(7.10) \quad \begin{aligned} \dot{\vec{y}} &= \begin{pmatrix} -500\frac{1}{2} & 499\frac{1}{2} \\ 499\frac{1}{2} & -500\frac{1}{2} \end{pmatrix} \vec{y} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ \vec{y}_0 &= \begin{pmatrix} -.1 \\ +.1 \end{pmatrix}. \end{aligned}$$

We hebben dit stelsel bij diverse staplengten twee maal geïntegreerd, één maal door uitsluitend formule (7.3) met $\mu = \mu(1000h)$ toe te passen, de andere maal door na één exponentiële aanpassing over te gaan op de trapezium regel. De resultaten zijn in tabel 7.2 vermeld.

Tabel 7.2

Afwijking van de exacte waarde in $t=1$ bij integratie van (7.10) met formule $F^{(1)}$.

h	$-10 \log \text{fout}$ exponentiële aanpassing	$-10 \log \text{fout}$ éénmaal exponentiële aanpassing $\mu = \frac{1}{2}$
.1	1.5	2.1
.05	1.8	2.7
.02	2.2	3.5
.01	2.5	4.2

Verder hebben we ons afgevraagd welke consequenties een foutieve schatting van de parameter μ heeft.

Beschouw de vergelijking

$$(7.11) \quad \begin{aligned} \dot{y} &= -\lambda y + 2, \\ y(0) &= y_0; \end{aligned}$$

Passen we hierop formule $F^{(1)}$ met $\mu = \mu(z_1)$ toe, dan vinden we de benadering

$$|y_1 - \tilde{y}(h)| \sim y_0 \frac{|z - z_1|}{z^2},$$

waarbij $z = h\lambda$. De numerieke resultaten komen met deze schatting overeen, getuige tabel 3.

Tabel 7.3

Fout bij integratie van (7.11) met $\lambda = 1000$ en $\mu = \mu(z_1)$.

z_1	200	110	101	100.1	100	99.9	99	90	10
$-10 \log y_1 - y(.1)$	2.3	3.0	4.0	5.0	11.0	5.0	4.0	2.9	1.0

De formules $F^{(2)}$ en $F^{(3)}$ vertonen een soortgelijk gedrag bij een onjuiste schatting van de parameters.

We kunnen dus op grond van nauwkeurigheidsoverwegingen de eis stellen

$$(7.12) \quad d\lambda < \lambda^2 \cdot h \cdot \text{tol},$$

waarbij $d\lambda$ de straal van de linker cluster, en tol de tolerantie in de stijve component van de oplossing is.

Tenslotte hebben we enkele strategieën getest bij het oplossen van het beginwaarde probleem uit de enzymkinetica (zie hoofdstuk 1).

$$(7.13) \quad \begin{aligned} \dot{s} &= (c-1)s = .99c, & s(0) &= 1, \\ \dot{c} &= -1000 ((c-1)s+c), & c(0) &= 0, \end{aligned}$$

te weten:

A1 : Eén iteratie, en bij iedere iteratie een Jacobiaan-evaluatie;

B1 : De eerste stap volgens A1, daarna één Jacobiaan-evaluatie per 5 stappen;

C1 : Als A1, met na 1 stap overgang op een hogere orde formule door in het punt $z_0 = 0$ exponentiëel te fitten;

D1 : Als B1, met na 1 stap fit in $z_0 = 0$;

A2, B2, C2 en D2 als A1, B1, C1 en D1, met dien verstande dat meermalen geïtereerd wordt (indien $\Delta y_{n+1}^{(1)} > \text{tol}$), met een maximum van 5.

Als integratiestap hebben we genomen $h = 5$, in de grootste negatieve eigenwaarde van de laatst berekende Jacobiaan hebben we $R(z)$ exponentiëel aangepast. De resultaten bij de formules $F^{(1)}$ en $F^{(3)}$ staan in de tabellen (7.4) en (7.5) vermeld.

Tabel 7.4

Fout in de numerieke oplossing, in $x = 50$, verkregen met formule $F^{(1)}$, bij diverse strategieën.

strategie	tol	$ s_{50} - \tilde{s}_{50} $	$ c_{50} - \tilde{c}_{50} $	a.f.e.	a.J.e.
A1		$1.8 \cdot 10^{-3}$	$6.5 \cdot 10^{-4}$	10	10
A2	$2 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$	$4.9 \cdot 10^{-4}$	21	21
A2	$2 \cdot 10^{-6}$	$1.5 \cdot 10^{-3}$	$4.9 \cdot 10^{-4}$	30	30
B1		$1.7 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	10	3
B2	$2 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$	$5.3 \cdot 10^{-4}$	21	6
B2	$2 \cdot 10^{-6}$	$1.5 \cdot 10^{-3}$	$4.9 \cdot 10^{-4}$	34	7
C1		$3.8 \cdot 10^{-4}$	$5.7 \cdot 10^{-3}$	10	10
C2	$2 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$4.5 \cdot 10^{-5}$	21	21
C2	$2 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	$4.5 \cdot 10^{-5}$	30	30
D1		$3.3 \cdot 10^{-4}$	$4.3 \cdot 10^{-3}$	10	3
D2	$2 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$6.0 \cdot 10^{-5}$	21	6
D2	$2 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	$4.5 \cdot 10^{-5}$	34	7

Tabel 7.5

Fout in de numerieke oplossing in $x = 50$, verkregen met formule $F^{(3)}$, bij diverse strategieën.

strategie	tol	$ s_{50} - \tilde{s}_{50} $	$ c_{50} - \tilde{c}_{50} $	a.f.e.	a.J.e.
A1		$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	10	10
A2	$2 \cdot 10^{-4}$	$1.8 \cdot 10^{-6}$	$5.9 \cdot 10^{-7}$	21	21
A2	$2 \cdot 10^{-6}$	$1.8 \cdot 10^{-6}$	$5.7 \cdot 10^{-7}$	30	30
B1		$7.2 \cdot 10^{-5}$	$6.2 \cdot 10^{-4}$	10	3
B2	$2 \cdot 10^{-4}$	$2.9 \cdot 10^{-5}$	$3.6 \cdot 10^{-5}$	21	5
B2	$2 \cdot 10^{-6}$	$3.3 \cdot 10^{-5}$	$1.1 \cdot 10^{-5}$	35	5
C1		$1.5 \cdot 10^{-4}$	$6.0 \cdot 10^{-3}$	10	10
C2	$2 \cdot 10^{-4}$	$1.6 \cdot 10^{-6}$	$5.7 \cdot 10^{-7}$	21	21
C2	$2 \cdot 10^{-6}$	$1.5 \cdot 10^{-6}$	$5.6 \cdot 10^{-7}$	30	30
D1		$1.3 \cdot 10^{-4}$	$3.4 \cdot 10^{-3}$	10	3
D2	$2 \cdot 10^{-4}$	$3.3 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	21	5
D2	$2 \cdot 10^{-6}$	$7.4 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$	35	5

De methode van Lawson (Gegeneraliseerde Runge-Kutta-processen)

Beschouw het m-dimensionale beginwaarde-probleem

$$(7.14) \quad \dot{y} = f(x, y), \quad y_0 = y(0).$$

We nemen aan dat dit een stijf stelsel is; de Jacobiaan J_f heeft dus eigenwaarden met sterk negatief reëel deel. Lawson stelt voor stelsel (7.14) te transformeren in een stelsel dat als oplossing heeft

$$(7.15) \quad z(x) = \exp(-xA) y(x),$$

waarin A een reële m x m matrix is.

Voor z(x) vinden we dan het beginwaarde-probleem

$$(7.16) \quad \begin{aligned} z'(x) &= g(x, z), \\ z_0 &= y_0 - \\ g(x, z) &= \exp(-xA) \{f[x, \exp(xA) z(x)] - A \exp(xA) z(x)\}. \end{aligned}$$

De Jacobiaan J_g van dit stelsel wordt gegeven door

$$J_g = \exp(-xA) \{J_f - A\} \exp(xA),$$

zodat de eigenwaarden van J_g dezelfde zijn als die van $(J_f - A)$. Stelsel (7.16) is dus niet stijf, indien A zo gekozen is dat de eigenwaarden van $(J_f - A)$ klein zijn. Hierbij is het dus niet noodzakelijk dat A gelijk is aan de Jacobiaan van (7.14).

Lawson stelt nu voor stelsel (7.16) op te lossen met een expliciete m-punts Runge-Kutta-methode van de vorm (3.10). Na substitutie van (7.15) in dit Runge-Kutta-schema ontstaat voor y(x) het schema

$$(7.17) \quad \left\{ \begin{array}{l} k_1^* = f(x_n, y_n) - Ay_n, \\ p_i^* = \exp(\mu_i h A) y_n + h \sum_{j=1}^{i-1} \lambda_{ij} \exp[(\mu_i - \mu_j) h A] k_j^*, \\ k_i^* = f(x_n + \mu_i h, p_i^*) - A p_i^*, \quad i = 2, 3, \dots, m, \\ y_{n+1} = \exp(h A) y_n + h \sum_{i=1}^m \theta_i \exp[(1 - \mu_i) h A] k_i^*, \end{array} \right.$$

$$n = 0, 1, 2, \dots$$

Bij dit schema dient een algorithmme gegeven te zijn, dat $\exp(\mu_i h A)$ berekend. De benadering met dit algorithmme geven we aan door $E(\mu_i h A)$. Lawson noemt als mogelijkheid hiervoor de diagonale Padé approximaties; deze vormen namelijk een verzameling onvoorwaardelijk stabiele benaderingen van de exponentiële functie. Verder moet een algorithmme gegeven zijn om de matrix A te bepalen. Het ligt voor de hand A gelijk te stellen aan de Jacobiaan; deze kan bij elke stap (algorithmme I) of periodiek (II) berekend worden. Lawson bewijst dat algorithmme I A -stabil is in de zin van Dahlquist, indien geldt

$$E(w) < 1 \quad \text{voor } \operatorname{Re}(w) < 0.$$

Over de globale fout e_n heeft Lawson de volgende stelling bewezen.

Stelling:

Zij gegeven een probleem van de vorm

$$(7.18) \quad \dot{y} = By + u(x, y), \quad y_0 = y(0),$$

waarbij u aan een Lipschitz-conditie voldoet,

$$|u(x, y) - u(x, z)| < L|y - z|.$$

Is bovendien voldaan aan de voorwaarden

$$u(x,0) = 0,$$

$$\lim_{x \rightarrow \infty} y(x) = 0$$

$$\text{en } ||E(hA)|| + h\alpha = \alpha < 1,$$

dan geldt
$$\lim_{n \rightarrow \infty} e_n = \lim_{n \rightarrow \infty} |y_n - y(x_n)| = 0.$$

(α is hierbij een functie van L en de Runge-Kutta parameters, zie Lawson).

Een nadeel van de methode van Lawson is, dat alle coëfficiënten van het stabiliteitspolynoom functies van de integratiestap h zijn. De stabiliteitsfunctie bij de methode van Euler na Lawson's transformatie wordt bijvoorbeeld gegeven door

$$R(z) = \beta_0 + \beta_1 z,$$

$$\beta_0 = e^{hA}(1-hA),$$

$$\beta_1 = e^{hA}.$$

Hieruit volgt dat de methode 1^e orde consistent is, want

$$\lim_{h \rightarrow 0} \beta_0 = \lim_{h \rightarrow 0} \beta_1 = 1$$

Bij een eindige staplengte zijn deze coëfficiënten echter ongelijk 1, en de afbreekfout bevat dus niet alleen de tweede en hogere afgeleiden.

Voorbeeld

Passen we de methode van Euler toe op de modelvergelijking

$$(7.19) \quad \dot{y} = -\delta y + x,$$

die als oplossing heeft $y(x) = \frac{x}{\delta} - \frac{1}{\delta^2}$, dan vinden we zonder transformatie

$$y_{n+1} = y(x_{n+1})$$

en met transformatie

$$y_{n+1} = (y_n + hx_n) e^{-h\delta}.$$

In het laatste geval is de afbreekfout (als $y_n = y(x_n)$)

$$y_{n+1} - y(x_{n+1}) = (1-\beta_0)y_n + (1-\beta_1) h \dot{y}_n;$$

we zien dus dat de afbreekfout inderdaad de 0^e en 1^e afgeleide van de oplossing bevat.

Numerieke resultaten

Lawson heeft een 4^e orde Runge-Kutta-methode met en zonder transformatie getest. De twee voorbeelden die hij hierbij koos waren stelsels van twee niet lineaire vergelijkingen. Voor de matrix A nam hij in beide gevallen de Jacobiaan van het lineaire gedeelte, voor E(A) werd de (2,2) Padé approximatie genoemd. Lawson's transformatie liet bij beide problemen een aanzienlijk grotere stap toe, terwijl de bereikte precisie van dezelfde orde van grootte was. Wel gold echter dat in beide problemen de asymptotische oplossing tot 0 naderde, dus $\lim_{x \rightarrow \infty} y(x) = 0$.

Wijzelf hebben Baanstra's 3^e orde Runge-Kutta-methode met automatisch stapkeuze mechanisme als basis genomen. De resultaten toegepast op stelsel (7.10) zijn opgenomen in tabel 7.6. Voor A kozen we de Jacobiaan; $\exp(hA)$ werd exact bepaald.

Tabel 7.6

Aantal functie evaluaties en fout bij integratie van (7.10) met 3^e orde Runge-Kutta

tol.	zonder transformatie		met transformatie	
	a.f.e.	$-10 \log y(1)-y_{\text{num}} $	a.f.e.	$-10 \log y(1)-y_{\text{num}} $
10^{-0}	1905	2.5	8	3.2
10^{-1}	1997	3.2	8	3.2
10^{-2}	2014	4.3	30	4.1
10^{-3}	2100	5.2	58	5.7
10^{-4}	2717	6.5	152	7.2
10^{-5}	4149	7.0	432	8.7

Tenslotte hebben we het probleem (7.13) uit de enzymkinetica opgelost met een 3^e orde Runge-Kutta formule met Lawson's transformatie. Als staplengte namen we $h = 5$. Om de nauwkeurigheid van de approximatie van $\exp(A)$ te vergroten, benaderden we $\exp(2^{-10}A)$ met een diagonale Padé approximatie. Daaruit volgde een benadering van $\exp(A)$, namelijk volgens de formule

$$[\exp(2^{-10}A)]^{2^{10}} = \exp(A).$$

We hebben het probleem met een aantal verschillende strategieën opgelost, en wel:

A : Jacobiaan-evaluatie en Padé-approximatie bij elke integratie-stap;

B : Jacobiaan-evaluatie en Padé-approximatie bij de 1^e en de 2^e integratie-stap
en verder om de 5 integratie-stappen;

C : Jacobiaan-evaluatie en Padé-approximatie bij de 1^e integratie-stap en verder periodiek met periode 5;

I : (1,1) - Padé-approximaties van $\exp(A)$;

II: (2,2) - Padé-approximaties van $\exp(A)$.

De resultaten staan vermeld in tabel 7.7.

Tabel 7.7.

Fout in de numerieke oplossing, in $x = 50$, verkregen met een 3^e orde Runge-Kutta-methode met Lawson's transformatie, en de aantallen functie- en Jacobiaan-evaluaties en Padé-approximaties.

strategie	$ s_{50} - \tilde{s}_{50} $	$ c_{50} - \tilde{c}_{50} $	a.f.e.	a.J.e.	a.P.a.
A I	$1.8 \cdot 10^{-3}$	$2.0 \cdot 10^{-1}$	30	2	2
B I	$2.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-1}$	30	3	3
C I	$5.4 \cdot 10^{-4}$	$1.3 \cdot 10^{-1}$	30	10	10
A II	$1.8 \cdot 10^{-3}$	$2.0 \cdot 10^{-1}$	30	2	2
B II	$2.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-1}$	30	3	3
C II	$5.4 \cdot 10^{-4}$	$1.3 \cdot 10^{-1}$	30	10	10

Het blijkt, dat de (1,1) en de (2,2) Padé-approximaties tot dezelfde resultaten leiden. In beide gevallen is de nauwkeurigheid in de tweede component gering.

Exponentiële aanpassing bij Runge-Kutta-methoden

Bij een expliciete Runge-Kutta-methode met een orde van consistentie p , kan een stabiliteitspolynoom worden opgesteld van de vorm

$$(7.20) \quad R_m(z) = 1 + z + \frac{1}{2!}z^2 + \dots + \frac{1}{p!}z^p + \beta_{p+1}z^{p+1} + \dots + \beta_m z^m.$$

In hoofdstuk 3 zijn reeds een aantal van deze polynomen vermeld; bovendien is een algoritme gegeven om bij een polynoom een expliciete p -de orde ($p \leq 3$) Runge-Kutta-methode te construeren, die dat polynoom tot stabiliteitsfunctie heeft.

In deze paragraaf gaan we nader in op exponentieel aangepaste polynomen. Hierbij beperken we ons tot polynomen van de vorm (7.20) met willekeurige p , waarbij $\beta_{p+1}, \dots, \beta_m$ nog vrij te kiezen parameters zijn. Deze vrijheidsgraden kunnen we benutten door $R_m(z)$ in een aantal punten tot een zekere orde aan de analytische stabiliteitsfunctie $\exp(z)$ aan te passen.

Indien we $R_m(z)$ in een complex punt z_0 exponentieel aanpassen, dan doen we dit ook in het geconjugeerde punt $\bar{z}_0$, teneinde complex rekenen te vermijden.

We verkrijgen op deze manier een stelsel lineaire vergelijkingen in de parameters $\beta_{p+1}, \dots, \beta_m$,

$$(7.21) \quad \begin{aligned} \frac{d^k}{dz^k} R_m(z_j) &= \exp(z_j), & k &= 0, 1, \dots, r_j, \\ j &= 1, 2, \dots, s. \end{aligned}$$

Hierin stelt r_j de multipliciteit van de aanpassing in het punt z_j voor. In het algemeen kan een oplossing van dit stelsel gevonden worden als

$$\sum_{j=1}^s r_j = m - p.$$

Stabiliteitsgebieden

We nemen aan dat de punten waarin de exponentiële aanpassing plaats vindt, alle een reëel deel veel kleiner dan 0 hebben. Dan kan het stabiliteits-

polynoom analoog aan Van der Houwen (1971) benaderd worden door

$$(7.22) \quad R_m(z) \sim (1+z+\dots + \frac{1}{p!} z^p) \prod_{j=1}^s \left(\frac{z-z_j}{z_j} \right)^{r_j}.$$

Uit deze representatie van $R_m(z)$ blijkt dat het stabiliteitsgebied in de omgeving van z_j bij benadering een cirkel is, waarvan de straal ρ_j gegeven wordt door

$$(7.23) \quad \rho_j^{r_j} = p! z_j^{(r_j-p)} \prod_{\substack{i=1 \\ i \neq j}}^s \left| \frac{z_i}{z_i - z_j} \right|^{r_i}.$$

Als speciale gevallen zullen we de twee- en drie-cluster methode beschouwen.

Twee- en drie-cluster methoden

Indien de stabiliteitsfunctie in slechts een punt z_0 exponentieel wordt aangepast, dan ontstaat de twee-cluster methode. Het blijkt dat de parameters $\beta_{p+1}, \dots, \beta_m$ dan expliciet als functie van z_0 gegeven kunnen worden. In Dekker (1972) staat hiervoor een algoritme, evenals voor de berekening van de parameters bij aanpassing in twee toegevoegd complexe punten (de drie-cluster methode), waarvoor evenwel een stelsel opgelost moet worden.

Uit formule (7.23) volgen de stabiliteitsgebieden van deze methoden.

In het geval van de twee-cluster methode vinden we

$$(7.24) \quad \rho_0 = z_0 \left| p! z_0^{-p} \right|^{\frac{1}{m-p}},$$

en bij de drie-cluster methode

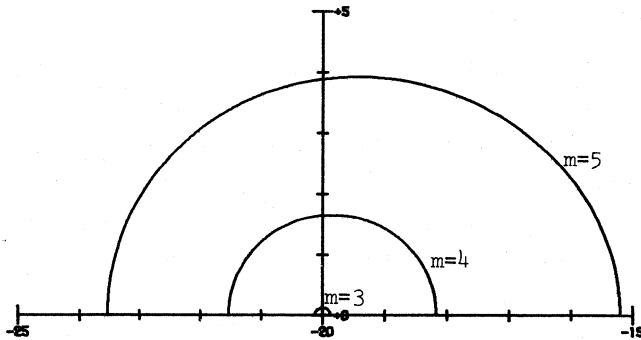
$$(7.25) \quad \rho_0 = z_0 \left| \frac{p! z_0^{-p}}{2 \sin \phi} \right|^{\frac{2}{(m-p)}};$$

hierbij is ϕ het argument van z_0 .

Uit deze formules voor ρ_0 volgt dat het stabiliteitsgebied bij z_0 des te groter wordt, naarmate de multipliciteit van de aanpassing toeneemt. Dit blijkt ook uit figuur 7.2, waarin we de linker stabiliteitsgebieden van de twee-cluster methode ($p=2$) geven. Meer afbeeldingen van stabiliteitsgebieden zijn in Dekker (1972) te vinden.

Figuur 7.2

Stabiliteitsgebieden bij $z_0 = -20$ van de twee-cluster methode, $p = 2$, $m = 3, 4, 5$.



Numerieke resultaten

Zoals hierboven reeds vermeld is, kunnen de Runge-Kutta-methoden van het type (3.72) als clustermethode gebruikt worden. Omdat in het algemeen de coëfficiënten van het stabiliteitspolynoom (deze zijn immers functies van z_j) per stap veranderen, moeten bij iedere stap de Runge-Kutta-parameters berekend worden.

Doordat echter de Runge-Kutta-parameters zeer eenvoudige functies van de coëfficiënten van $R_m(z)$ zijn, kost dit weinig extra rekenwerk.

In Dekker (1972) staan een aantal resultaten met de twee- en drie-cluster methode vermeld. Daarbij bleek dat bij een niet-lineair probleem met snel variërende Jacobiaan een lage orde formule met een hoge multipliciteit in de aanpassing de beste resultaten gaven. Bij lineaire problemen was een 1^e orde aanpassing voldoende om met grote staplengten te kunnen integreren. Verder bleek dat de 3^e orde formules bij grote integratie stappen door interne instabiliteit verstoord werden. Dit verschijnsel trad ook op bij het oplossen van probleem (7.13) met de twee-cluster methode.

De resultaten staan in tabel 7.8 vermeld.

Tabel 7.8

Fout in de numerieke oplossing van (7.13) in $x = 50$
verkregen met de twee-cluster methode.

r	m	aantal stappen	$-^{10}\log s_{50}-\tilde{s}_{50} $	$-^{10}\log c_{50}-\tilde{c}_{50} $
1	2,3,4,5,6	101	3.8	4.3
1	3,4,5,6	51	3.5	4.0
1	3,4,5,6	26	3.2	3.7
1	3,4,5,6	14	2.9	3.4
1	4,5,6	8	2.6	3.1
1	4,5,6	5	2.3	2.9
2	4,5,6,7	104	7.2	4.5
2	5,6,7	55	5.6	2.6
2	7	39	6.0	3.1

Literatuur

- 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.
- Houwen, P.J. van der A survey of stabilized Runge-Kutta methods,
MC Tract 37, Ch. 5, Mathematisch Centrum,
Amsterdam (1971).
- Dekker, K. Een ALGOL 60-versie van exponentieel aangepaste Runge-Kutta methoden.
NR 25, Mathematisch Centrum, Amsterdam (1972).

8. Semi-Runge-Kutta-methoden

Zij gegeven de differentiaalvergelijking

$$(8.1) \quad \frac{dy}{dx} = f(y)$$

met beginvoorwaarde $y_0 = y(x_0)$ en waarin $f(y)$ een vectorfunctie is die voldoende vaak naar y differentieerbaar is.

In hoofdstuk 3 werd een integratietechniek ingevoerd welke gebaseerd was op herhaalde evaluatie van het rechterlid en de Jacobiaan (bijvoorbeeld de formules van Rosenbrock (3.18) en Calahan (3.19) en de MC-formules (3.82) en (3.84)). Deze methode werd een semi-Runge-Kutta-methode genoemd en had de algemene vorm

$$(8.2) \quad \begin{aligned} y_{n+1} &= y_n + \sum_{j=0}^{m-1} \theta_j k_n^{(j)}, \\ k_n^{(j)} &= h_n f(y_n + \sum_{l=0}^{j-1} \Lambda_{j,l} k_n^{(l)}), \quad j = 0, 1, \dots, m-1, \\ \Lambda_{0,0} &= \Lambda_{0,-1} = 0. \end{aligned}$$

Hierin is y_n de numerieke benadering van $y(x)$ in $x = x_n$, h_n de $n+1$ -integratiestap $x_{n+1} - x_n$, en θ_j en $\Lambda_{j,l}$ zijn functies van $h_n J_n$ met J_n de Jacobiaan van het rechterlid in $y = y_n$.

Als in (8.2) één of meer van de functies θ_j of $\Lambda_{j,l}$ een rationale functie is, noemen we de methode *semi-impliciet*, terwijl de methode *expliciet* is als alle θ_j en $\Lambda_{j,l}$ polynomen zijn.

De integratiemethoden, die in dit hoofdstuk gegeven worden, zijn van type (8.2) en voldoen aan de volgende eisen

- a. per stap weinig functie-evaluaties en slechts één Jacobiaan-evaluatie;
- b. een redelijke orde van nauwkeurigheid;
- c. A-stabiliteit of een stabiliteitsgebied bestaande uit een gebied bij de oorsprong en één of meer gebieden in het linkerhalfvlak ver verwijderd van de oorsprong (twee- en drie-clustermethoden);

- d. de mogelijkheid de stabiliteitsfunctie te identificeren met een voorgeschreven stabiliteitsfunctie;
- e. exponentiële aanpassing;
- f. bij semi-impliciete methoden hoogstens één L-U ontbinding per stap.

We zullen nu in het kort aangeven aan welke eisen de bovengenoemde formules (3.18), (3.19), (3.82) en (3.84) niet voldoen.

Rosenbrock-formule (3.18) en Calahan-formule (3.19)

Deze methoden worden gedefinieerd door

$$\begin{aligned}
 \theta_0 &= 0, \\
 (8.3) \quad \theta_1 &= \frac{1}{1 - (1 - \frac{1}{2} \sqrt{2}) h_n J_n}, \\
 \Lambda_{1,0} &= \frac{1}{2} (\sqrt{2} - 1) \theta_1,
 \end{aligned}$$

(zie Rosenbrock (1963))

$$\begin{aligned}
 \text{en} \quad \theta_0 &= \frac{3}{4(1 - \frac{1}{2}(1 + \frac{1}{3} \sqrt{3}) h_n J_n)}, \\
 (8.4) \quad \theta_1 &= \frac{1}{3} \theta_0, \\
 \Lambda_{1,0} &= -\frac{8}{9} \sqrt{3} \theta_0,
 \end{aligned}$$

(zie Calahan (1968)).

De Rosenbrock-formule heeft het nadeel dat de methode slechts 2^e orde nauwkeurig is en niet exponentieel kan worden aangepast. De Calahan-formule is weliswaar 3^e orde consistent maar kan eveneens niet exponentieel worden aangepast.

Formules van Liniger en Willoughby

In hoofdstuk 7 werden behandeld de *impliciete* formules van Liniger en

Willoughby ($F^{(1)}$, $F^{(2)}$ en $F^{(3)}$). Lost men deze op met Newton-iteratie maar beperkt men zich tot één iteratie-slag dan behoren deze formules tot de klasse van integratiemethoden (8.2) met $m = 1$ en respectievelijk

$$(8.5) \quad \theta_0 = \frac{1}{1 - (1-\mu)h_n J_n},$$

$$(8.6) \quad \theta_0 = \frac{1 - \frac{1}{2}ah_n J_n}{1 - \frac{1}{2}h_n(1+a)J_n + \frac{1}{4}h_n^2(b+a)J_n^2},$$

$$(8.7) \quad \theta_0 = \frac{1 - \frac{1}{2}a h_n J_n}{1 - \frac{1}{2}h_n(1+a)J_n + \frac{1}{12}h_n^2(1+3a)J_n^2}.$$

De formules corresponderend met (8.5) en (8.6) zijn respectievelijk 1^e en 2^e orde exact maar de formule behorend bij (8.7) is in tegenstelling tot de $F^{(3)}$ formule slechts 2^e orde exact.

De MC-formules (3.82) en (3.84)

Deze formules zijn gedefinieerd door

$$(8.8) \quad \begin{aligned} \theta_0 &= \frac{3\alpha^2-1}{3\alpha^2} \left(1 + \frac{1}{2}\alpha \frac{3\alpha-2}{3\alpha^2-1} h_n J_n + \frac{1}{2}\alpha^2 \frac{3\alpha^3-3\alpha+1}{(3\alpha^2-1)^2} h_n^2 J_n^2\right), \\ \theta_1 &= \frac{1}{3\alpha^2}, \\ \Lambda_{1,0} &= \frac{3\alpha^3}{3\alpha^2-1} \theta_0, \end{aligned}$$

en

$$(8.9) \quad \begin{aligned} \theta_0 &= \frac{3\alpha^2-1}{3\alpha^2} \frac{1 - \frac{1}{12} \frac{1+3\alpha_1}{\alpha} h_n J_n}{1 - \frac{1}{2}(1+\alpha_1)h_n J_n + \frac{1}{12}(1+3\alpha_1)h_n^2 J_n^2}, \\ \theta_1 &= \frac{1}{3\alpha^2}, \\ \Lambda_{1,0} &= \frac{3\alpha^3}{3\alpha^2-1} \theta_0. \end{aligned}$$

Hierin wordt de parameter α bepaald door de vergelijkingen (3.83) respectievelijk (3.86).

Deze methoden zijn 3^e orde exact en kunnen exponentieel worden aangepast maar ze hebben het nadeel dat ze moeilijk in een zodanige vorm kunnen worden gebracht dat de stabiliteitsfunctie van de methode geïdentificeerd kan worden met een voorgeschreven stabiliteitsfunctie. Hiervoor zou namelijk een niet lineair stelsel vergelijkingen opgelost moeten worden (zie van der Houwen (1972a)).

Constructie van 3^e orde formules met voorgeschreven stabiliteitsfunctie

We geven nu een afleiding van tweepuntsformules die voldoen aan de eisen a - f. Nemen we in (8.2) $m = 2$, dan krijgen we

$$(8.10) \quad y_{n+1} = y_n + \theta_0(z) h_n f(y_n) + \theta_1(z) h_n f(y_n + \Lambda_{1,0}(z) h_n f(y_n)).$$

In hoofdstuk 3 is al afgeleid dat (8.10) 1^e orde nauwkeurig is als de coëfficiënten voldoen aan

$$(8.11) \quad \theta_0(0) + \theta_1(0) = 1,$$

en 2^e orde nauwkeurig als bovendien

$$(8.12) \quad \theta_0'(0) + \theta_1'(0) + \theta_1(0) \Lambda_{1,0}'(0) = \frac{1}{2},$$

terwijl voor 3^e orde consistentie nog voldaan moet worden aan de voorwaarden

$$(8.13) \quad \begin{aligned} \theta_0''(0) + \theta_1''(0) + 2\theta_1(0) \Lambda_{1,0}'(0) + 2\theta_1'(0) \Lambda_{1,0}(0) &= \frac{1}{3}, \\ \theta_1(0) \Lambda_{1,0}^2(0) &= \frac{1}{3}. \end{aligned}$$

Uit deze laatste voorwaarde blijkt dat éénpuntsformules zoals (8.5), (8.6) en (8.7) niet 3^e orde nauwkeurig kunnen zijn.

Vervolgens zullen we trachten de stabiliteitsfunctie van (8.10) te identificeren met een voorgeschreven stabiliteitsfunctie, die om de nauwkeurigheid niet te verstoren 3^e orde consistent moet zijn.

Passen we (8.10) toe op de modelvergelijking $y' = \delta y$, met δ een willekeurig complex getal, dan vinden we voor de stabiliteitsfunctie $R(z)$ de uitdrukking

$$(8.14) \quad R(z) = 1 + z [\theta_0(z) + \theta_1(z)] + z^2 \theta_1(z) \Lambda_{1,0}(z).$$

Het blijkt mogelijk te zijn aan de consistentie-voorwaarden te voldoen binnen de klasse van formules waarvoor θ_0 en θ_1 constanten zijn, i.e.

$$(8.15) \quad \theta_0(z) \equiv \theta_0,$$

$$(8.16) \quad \theta_1(z) \equiv \theta_1.$$

(Zie van der Houwen (1972b)).

We vinden nu

$$(8.17) \quad \theta_0 = \frac{1}{4} \text{ en } \theta_1 = \frac{3}{4}.$$

Voor een willekeurige stabiliteitsfunctie

$$(8.18) \quad R(z) = \frac{\sum_{j=0}^{m_2} \beta_j z^j}{\sum_{j=0}^{m_1} \alpha_j z^j}, \quad \alpha_0 = 1,$$

blijkt uit tabel 3.1 3^e orde consistentie voor

$$(8.19) \quad \begin{aligned} \beta_0 &= 1, \\ \beta_1 &= 1 + \alpha_1, \\ \beta_2 &= \frac{1}{2} + \alpha_1 + \alpha_2, \\ \beta_3 &= \frac{1}{6} + \frac{1}{2}\alpha_1 + \alpha_2 + \alpha_3. \end{aligned}$$

Met behulp van deze voorwaarden vinden we nu voor $m_1 \geq m_2$

$$\begin{aligned}
 \Lambda_{1,0}(z) &= \\
 (8.20) \quad & 1 + \frac{1}{3}(1+3\alpha_1)z + 2 \sum_{j=4}^{m_2} (\beta_j - \alpha_j - \alpha_{j-1})z^{j-2} + 2 \sum_{j=m_2+1}^{m_1} (-\alpha_j - \alpha_{j-1})z^{j-2} - 2\alpha_{m_1} z^{m_1-1} \\
 &= \frac{2}{3} \cdot \frac{1 + \frac{1}{3}(1+3\alpha_1)z + 2 \sum_{j=4}^{m_2} (\beta_j - \alpha_j - \alpha_{j-1})z^{j-2} + 2 \sum_{j=m_2+1}^{m_1} (-\alpha_j - \alpha_{j-1})z^{j-2} - 2\alpha_{m_1} z^{m_1-1}}{1 + \alpha_1 z + \dots + \alpha_{m_1} z^{m_1}}.
 \end{aligned}$$

We hebben nu afgeleid een klasse van 3^e orde integratiemethoden waarvan de stabiliteitsfunctie vrij gekozen kan worden mits die compatibel is met de orde van de methode.

De procedure efsirk (exponential fitted semi-implicit Runge-Kutta)

Als speciaal geval van de nu ontwikkelde klasse van 3^e orde twee-punts methoden nemen we de stabiliteitsfunctie van de $F^{(3)}$ formule van Liniger en Willoughby:

$$(8.21) \quad R(z) = \frac{1 + \frac{1}{2}(1-\alpha_1)z + \frac{1}{12}(1-3\alpha_1)z^2}{1 - \frac{1}{2}(1+\alpha_1)z + \frac{1}{12}(1+3\alpha_1)z^2}.$$

Met deze stabiliteitsfunctie krijgen we nu voor (8.20) de uitdrukking

$$(8.22) \quad \Lambda_{1,0}(z) = \frac{2}{3} \frac{1 - \frac{1}{6}(1+3\alpha_1)z}{1 - \frac{1}{2}(1+\alpha_1)z + \frac{1}{12}(1+3\alpha_1)z^2}.$$

We hebben nu dus gekregen de semi-impliciete twee-puntsformule

$$\begin{aligned}
 (8.23) \quad y_{n+1} &= y_n + \frac{1}{4}h_n f(y_n) \\
 &+ \frac{1}{3}h_n f(y_n) + \frac{2}{3} \frac{1 - \frac{1}{6}(1+3\alpha_1)h_n J_n}{1 - \frac{1}{2}(1+\alpha_1)h_n J_n + \frac{1}{12}(1+3\alpha_1)h_n^2 J_n^2} h_n f(y_n),
 \end{aligned}$$

waarin α_1 een vrije parameter is, die gebruikt kan worden om de methode

exponentieel aan te passen in $h_n \delta$, met δ de in absolute waarde grootste eigenwaarde van de Jacobiaan. (Om complex rekenen te vermijden nemen we in de praktijk δ reëel of δ gelijk aan het reële gedeelte van een complexe eigenwaarde.) Voor $h_n \delta \leq 0$ geldt $0 \leq \alpha_1 \leq \frac{1}{3}$ en Liniger en Willoughby bewezen al dat de stabiliteitsfunctie A-stabiel is voor $\alpha_1 \geq 0$.

Testvoorbeelden

Een lineair stelsel van Fowler en Warten

We passen de nu ontwikkelde methode toe op het stelsel differentiaalvergelijkingen (Fowler en Warten (1967))

$$(8.24) \quad y' = \begin{pmatrix} -500.5 & 499.5 \\ 499.5 & -500.5 \end{pmatrix} y + \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad y(0) = \begin{pmatrix} -0.1 \\ 0.1 \end{pmatrix}.$$

De analytische oplossing van dit stelsel luidt

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

De Jacobiaan van het stelsel heeft de eigenwaarden $\delta_1 = -1$ en $\delta_2 = -1000$. Probleem (8.24) is opgelost op het interval $[0,10]$ door integratie met een constante stap. De methode is exponentieel aangepast in -1000 , -10^{100} en 0 .

In tabel 8.1, 8.2 en 8.3 is voor de verschillende waarden waarin de methode is aangepast een overzicht gegeven van het aantal significante cijfers:

$$-10_{\log} \left[\max_{k=1,2} \left| \frac{y^{(k)}(x_n) - y_n^{(k)}}{y^{(k)}(x_n)} \right| \right],$$

in $x_n = 1, 2$ en 10 .

Tabel 8.1

- $10^{\log}$ rel. fout voor probleem (8.24)
aangepast in -1000.

x	Staplenge						
	10	5	1	0.5	0.1	0.05	0.01
1			2.2	3.0	5.1	6.0	8.3
2			2.4	3.3	5.4	6.3	8.6
10	1.0	2.2	5.3	6.2	8.2	9.1	11.7

Tabel 8.2

- $10^{\log}$ rel. fout voor probleem (8.24)
aangepast in -10^{100} .

x	Staplenge						
	10	5	1	0.5	0.1	0.05	0.01
1			2.2	3.0	5.1	6.0	8.1
2			2.4	3.3	5.4	6.3	8.4
10	1.0	2.2	5.3	6.2	8.2	9.1	11.7

Tabel 8.3

- $10^{\log}$ rel. fout voor probleem (8.24)
aangepast in 0.

x	Staplenge						
	10	5	1	0.5	0.1	0.05	0.01
1			1.1	1.1	1.6	3.2	10.6
2			1.2	1.3	2.3	5.4	11.4
10	0.5	1.2	1.4	1.5	6.5	10.4	11.7

Uit tabel 8.1 en 8.2 blijkt dat het de beste strategie is de methode exponentieel aan te passen in de kleinste eigenwaarde, maar dat de methode niet gevoelig is voor een foute schatting van deze eigenwaarde.

In tabel 8.3 zien we dat aanpassing in 0 voor kleine integratiestappen goede resultaten geeft. Dit is een gevolg van de 4^e orde consistentie van de methode voor lineaire differentiaalvergelijkingen, als de methode in 0 exponentieel is aangepast.

De vergelijking van Robertson

Als tweede voorbeeld nemen we het stelsel vergelijkingen (Robertson (1967) en Lapidus & Seinfeld (1971))

$$(8.25) \quad \begin{aligned} y_1' &= -0.04 y_1 + 10^4 y_2 y_3, \\ y_2' &= 0.04 y_1 - 10^4 y_2 y_3 - 3_{10} 7 y_2^2, \\ y_3' &= 3_{10} 7 y_2^2, \end{aligned} \quad y(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Deze vergelijkingen zijn afhankelijk en met behulp van de beginvoorwaarde krijgen we

$$y_1(x) + y_2(x) + y_3(x) = 1.$$

De oplossing gaat asymptotisch naar $y_1 = 0$, $y_2 = 0$ en $y_3 = 1$. In $x = 0$ vinden we voor de eigenwaarden $\delta_1 = \delta_2 = 0$ en $\delta_3 = -0.04$, terwijl voor $x > 0$, $\delta_1 = 0$, $\delta_2 \approx 0$ en $\delta_3 \rightarrow -10^4$. Dit probleem blijkt dus een zeer stijf karakter te hebben. Hier komt nog bij dat de eigenwaarde aanvankelijk zeer snel afneemt. Omdat van (8.25) geen analytische oplossing bekend is, hebben we een referentie-oplossing berekend met een kleine staplengte, namelijk:

$$\begin{aligned} h_n &= 0.001, \quad 0 \leq x_n < 0.4, \\ h_n &= 0.005, \quad 0.4 \leq x \leq 10. \end{aligned}$$

De waarden van deze referentie-oplossing in $x = 0.4$ en $x = 10$ staan in tabel 8.4.

Tabel 8.4

Referentie-oplossing voor probleem (8.25).

x	y_1	y_2	y_3
0.4	.98517 21132	.33863 95382 ₁₀ ⁻⁴	.14794 02281 ₁₀ ⁻¹
10	.84136 99237	.16233 91063 ₁₀ ⁻⁴	.15861 38424

Integreren we probleem (8.25) met een kleine constante staplengte $h_n = 0.01$, dan krijgen we voor bv. y_3 de resultaten gegeven in tabel 8.5.

Tabel 8.5

Berekende waarde van y_3
na n stappen met $h_n = 0.01$.

n	y_3
1	.16 ₁₀ ⁻¹
2	.24 ₁₀ ²
3	.59 ₁₀ ⁸
4	.35 ₁₀ ²¹
5	.12 ₁₀ ⁴⁷

Dit instabiele gedrag van de oplossing wordt veroorzaakt door het sterk niet-lineaire gedrag in de eerste stappen. Hierdoor wordt gesuggereerd het probleem op te lossen met aanvankelijk zeer kleine stappen. In tabel 8.6 staat de absolute fout van de oplossing verkregen met

$$h_n = 0.001 \quad , \quad n < 4,$$

$$h_n = 0.1 \quad , \quad n \geq 4.$$

Tabel 8.6

Absolute fout voor probleem (8.25)

met $h_n = 0.001$, $n < 4$, $h_n = 0.1$, $n \geq 4$.

x	y_1	y_2	y_3
0.4	$.2_{10}^{-6}$	$.2_{10}^{-6}$	$.9_{10}^{-8}$
10	$.8_{10}^{-6}$	$.2_{10}^{-7}$	$.8_{10}^{-6}$

Stapkeuze strategie

Uit het laatste voorbeeld blijkt dat de grootte van de integratiestap afhankelijk is van de mate waarin de differentiaalvergelijking afwijkt van een lineaire differentiaalvergelijking. Om de procedure zelf de grootte van de stap te laten bepalen berekenen we een referentie-oplossing $\tilde{y}_n$ met dezelfde stabiliteitsfunctie als de methode. Immers in dat geval is het verschil tussen $\tilde{y}_n$ en y_n afhankelijk van het niet-lineaire gedrag van de vergelijking. Wij zoeken de referentie-oplossing in de vorm

$$(8.26) \quad \begin{aligned} \tilde{y}_{n+1} = y_n &+ v_0 h_n f(y_n) + v_1 h_n f(y_n + \Lambda_{1,0} h_n f(y_n)) \\ &+ v_2 h_n \Lambda_{1,0} h_n f(y_n) + v_3 h_n f(y_{n+1}). \end{aligned}$$

Hiervan is de stabiliteitsfunctie

$$(8.27) \quad \tilde{R}(z) = 1 + v_0 z + v_1 z(1 + \Lambda_{1,0} z) + v_2 \Lambda_{1,0} z^2 + v_3 z R(z).$$

Veronderstellen we nu dat $\Lambda_{1,0}$ gegeven wordt door (8.22) en $R(z)$ door (8.21) en eisen we dat $\tilde{R}(z) = R(z)$ dan moeten de coëfficiënten voldoen aan de vergelijkingen

$$\begin{aligned}
 & 3v_0 + 3v_1 + 2v_2 + 3v_3 = 3, \\
 (8.28) \quad & 9(1+\alpha_1)v_0 - 3(1-3\alpha_1)v_1 + 2(1+3\alpha_1)v_2 - 9(1-\alpha_1)v_3 = 9\alpha_1, \\
 & -3(1+3\alpha_1)v_0 + (1+3\alpha_1)v_1 + (9\alpha_1-3)v_3 = 0.
 \end{aligned}$$

Stellen we nu $v_0 = 0$ dan vinden we als oplossing

$$\begin{aligned}
 & v_0 = 0, \\
 (8.29) \quad & v_1 = \frac{1}{2} - 2v_3, \\
 & v_2 = \frac{1}{4}(1+2v_3), \\
 & v_3 = \frac{1}{2} \frac{1+3\alpha_1}{5-3\alpha_1}.
 \end{aligned}$$

Omdat de benadering (8.26) 2^e orde exact is geldt

$$(8.30) \quad \|\tilde{y}_{n+1} - y_{n+1}\| = c_n h_n^3.$$

Veronderstellen we nu dat

$c_n \approx c_{n-1}$, dan vinden we

$$(8.31) \quad h_n = \sqrt[3]{\frac{\text{tol}}{c_n}} \approx \sqrt[3]{\frac{\text{tol}}{c_{n-1}}} = h_{n-1} \sqrt[3]{\frac{\text{tol}}{\|\tilde{y}_n - y_n\|}},$$

waarin tol een door de gebruiker op te geven lokale tolerantie is. In de procedure efsirk is voor (8.31) de benadering gebruikt

$$(8.32) \quad h_n = \left(\frac{\frac{4}{3} \text{tol}}{\text{tol} + \|\tilde{y}_n - y_n\|} + \frac{1}{3} \right) h_{n-1}.$$

Voor de lokale tolerantie is in efsirk genomen

$$\text{abs. } \epsilon + \text{rel. } \epsilon \|y_n\|_2,$$

waarin $\|\cdot\|_2$ de Euclidische norm is.

Met deze stapkeuze strategie is probleem (8.25) opgelost met $\text{abs. } \epsilon = \text{rel. } \epsilon = 10^{-6}$ en met $\text{abs. } \epsilon = \text{rel. } \epsilon = 10^{-4}$. Voor de minimale staplengte hebben we genomen $h = 0.001$ en voor de maximale staplengte $h = 0.1$ voor $x < 5$ en $h = 0.2$ voor $5 \leq x \leq 10$. In tabel 8.7 zijn de absolute fout van de componenten en het aantal stappen gegeven.

Tabel 8.7

Absolute fout en het aantal stappen voor probleem (8.25).

x	abs. $\epsilon = \text{rel. } \epsilon = 10^{-6}$				abs. $\epsilon = \text{rel. } \epsilon = 10^{-4}$			
	y_1	y_2	y_3	n	y_1	y_2	y_3	n
0.4	$.4 \cdot 10^{-7}$	$.3 \cdot 10^{-8}$	$.4 \cdot 10^{-7}$	17	$.2 \cdot 10^{-6}$	$.4 \cdot 10^{-7}$	$.2 \cdot 10^{-6}$	13
10	$.3 \cdot 10^{-6}$	$.1 \cdot 10^{-9}$	$.3 \cdot 10^{-6}$	134	$.3 \cdot 10^{-5}$	$.1 \cdot 10^{-7}$	$.3 \cdot 10^{-5}$	86

Een niet-autonome vergelijking

Als derde voorbeeld nemen we de differentiaalvergelijking (Lapidus & Seinfeld (1971))

$$(8.33) \quad y' = -200(y - F(x)) + F'(x), \quad y(0) = 10,$$

$$F(x) = 10 - (10+x)e^{-x}.$$

De analytische oplossing van probleem (8.33) luidt

$$y(x) = F(x) + 10 e^{-200x},$$

en de kleinste eigenwaarde is -200 . Het probleem is opgelost op het interval $[0, 10]$ met een variërende stapgrootte. In tabel 8.8 staan het aantal significante cijfers en het aantal stappen in $x = 0.4$ en $x = 10$ voor $\text{abs. } \epsilon = \text{rel. } \epsilon = 10^{-2}$ en voor $\text{abs. } \epsilon = \text{rel. } \epsilon = 10^{-5}$. Voor de minimale staplengte is genomen $h = 0.005$ en voor de maximale stap $h = 0.5$.

Tabel 8.8

$-^{10}\log.$ rel. fout en het aantal stappen voor probleem (8.33).

x	abs. ϵ = rel. ϵ = 10^{-2}	n	abs. ϵ = rel. ϵ = 10^{-5}	n
0.4	1.6	13	2.3	64
10	3.4	86	5.9	449

Een lineair stelsel met complexe eigenwaarden

Op het colloquium (zie p. 58) is al ter sprake gekomen de differentiaal-vergelijking

$$(8.34) \quad y' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10^6 & -10^6 - 10^3 & -10^3 - 1 \end{pmatrix} y, \quad y(0) = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix},$$

met de analytische oplossing

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

De eigenwaarden van de Jacobiaan zijn

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

Het reële gedeelte van δ_2 en δ_3 is -500 en de modulus 1000. De differentiaal-vergelijking is opgelost op het interval $[0,1]$ met exponentiële aanpassing in -500 en in -1000. In tabel 8.9 wordt het aantal juiste cijfers gegeven in $x = 0.5$ en $x = 1$.

Tabel 8.9

- $10 \log \text{rel. fout}$ voor probleem (8.34).

x	$\delta = -500$					$\delta = -1000$				
	Staplengthte					Staplengthte				
	1	0.5	0.1	0.05	0.01	1	0.5	0.1	0.05	0.01
0.5		3.1	5.2	6.1	8.4		3.1	5.2	6.1	8.3
1	1.9	2.8	4.9	5.8	8.1	1.9	2.8	4.9	5.8	7.9

De procedure eferk (exponential fitted explicit Runge Kutta)

Als een tweede speciaal geval van de klasse van derde orde twee-punts methoden beschouwen we methoden waarvan de functie $\Lambda_{1,0}(z)$ een polynoom in z is (vergelijk formule 8.20).

Hiermede wordt een klasse van expliciete methoden verkregen van de vorm

$$(8.35) \quad y_{n+1} = y_n + \frac{1}{4}h_n f(y_n) + \frac{3}{4}h_n f\left\{y_n + \frac{4}{3}\left(\frac{1}{2} + \frac{1}{6}h_n J_n + \sum_{j=4}^{m_2} \beta_j (h_n J_n)^{j-2}\right)h_n f(y_n)\right\}.$$

De stabiliteitsfunctie van deze formule is het polynoom

$$(8.36) \quad R(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \sum_{j=4}^{m_2} \beta_j z^j.$$

De coëfficiënten β_j , $j = 4, \dots, m_2$, zijn hierbij vrije parameters, die zo gekozen kunnen worden, dat aanpassing van $R(z)$ aan de exponentiële operator in een aantal punten $z_0, z_1, \dots, z_{m_2-4}$ bereikt wordt (vergelijk hoofdstuk 7, pagina 170 e.v.).

Evenals in hoofdstuk 7 beperken we ons tot twee- en drie-cluster-methoden, die ontstaan door exponentiële aanpassing van $R(z)$ in één reëel, respectievelijk twee complex geconjugeerde punten.

Testvoorbeelden

Voor lineaire differentiaal-vergelijkingen, waarbij de eigenwaarden van

de Jacobiaan ver uiteen liggen, zijn deze expliciete methoden zeer geschikt. Terwijl een redelijke orde van nauwkeurigheid bereikt wordt en de mogelijkheid van exponentiële aanpassing bestaat, vergen de methoden een relatief geringe hoeveelheid rekenwerk in vergelijking met impliciete of semi-impliciete methoden. We illustreren dit met tabel (8.10), waarin de resultaten vermeld zijn van integratie met constante stap van probleem (8.24).

Tabel 8.10

$-10 \log \text{rel.fout}$ voor probleem (8.24); aangepast in -1000 .

x	m_2	Staplenkte						
		10	5	1	.5	.1	.05	.01
1	4			1.3	2.3	4.6	5.5	7.8
1	7			1.3	2.4	4.7	5.6	8.3
2	4			1.6	2.6	4.9	5.8	8.0
2	7			1.6	2.6	4.9	5.9	8.6
10	4	-2.1	-2.2	4.5	5.5	7.7	8.6	11.4

We merken hierbij op, dat voor staplengten $h > 2.5$ het produkt van h en de kleine eigenwaarde (namelijk -1) niet in het stabiliteitsgebied in de omgeving van de oorsprong ligt. Bij deze staplengten ontstaat daardoor een aanzienlijke fout.

Omdat instabiliteit kan optreden, indien de stabiliteitsfunctie niet in het juiste punt wordt aangepast, is het noodzakelijk om de grootste eigenwaarde van de Jacobiaan nauwkeurig te bepalen. De stabiliteitsgebieden zijn namelijk gering van omvang, zoals uit de formules (7.24) en (7.25) volgt. Uit deze formules kan gemakkelijk de maximale staplenkte die op grond van stabiliteit genomen mag worden, worden afgeleid, indien het verschil $\delta\delta_0$ tussen de kleinste eigenwaarde en het punt δ_0 waarin aangepast wordt, gegeven is. Deze maximale staplenkte luidt

$$(8.37) \quad h_{\text{stab}} = \frac{1}{\delta_0} \sqrt[3]{6} \left(\frac{\delta_0}{d\delta_0} \right)^{\frac{m_2}{3}} - 1.$$

Uit deze formule blijkt dat deze maximale staplengte groter wordt, indien m_2 toeneemt.

We hebben probleem (8.24) nogmaals geïntegreerd, maar nu met aanpassing in het punt -1010. De resultaten bij diverse waarden van m_2 en diverse staplengten staan vermeld in tabel 8.11. Bovendien wordt hierin de met formule (8.37) berekende maximaal toegelaten stap gegeven.

Tabel 8.11

$-^{10}\log \text{rel.fout in } x=1 \text{ voor probleem (8.24); aangepast in -1010.}$

m_2	hstab	staplengte				
		1	.5	.1	.05	.01
4	$8.4_{10^{-3}}$	-5.1	-9.5	-13.0	-13.0	4.7
5	$3.9_{10^{-2}}$	-3.1	-5.5	-10.6	-3.5	7.9
6	$1.8_{10^{-1}}$	-1.1	-1.5	4.6	5.6	8.1
7	$8.4_{10^{-1}}$	.7	2.1	4.7	5.6	8.3
8	3.9	1.3	2.4	4.7	5.7	8.5
9	18	1.3	2.4	4.7	5.7	8.7

Uit het bovenstaande mogen we concluderen dat deze methoden, met constante staplengte uitgevoerd, minder goede resultaten zullen opleveren bij het oplossen van (sterk) niet-lineaire problemen; de eigenwaarden zullen dan immers veranderen tijdens het integratie-proces. Een voorbeeld van een niet-lineair probleem is het stelsel (8.25). We hebben getracht dit stelsel met een integratie-stap van 10^{-2} te integreren; de verkregen resultaten geven we in tabel 8.12.

Tabel 8.12

Berekende waarde van y_3 na
n stappen met $h_n = 0.01$.

n	y_3	
1	1.6	10^{-2}
2	1.6	10^{22}
3	2.3	10^{167}
4	1.3	10^{626}
5	3.1	10^{626}

Integreren we dit probleem aanvankelijk met een kleine staplengte, dan verkrijgen we de resultaten die in tabel 8.13 vermeld staan.

Tabel 8.13

Absolute fout voor probleem (8.25) met
 $h_n = 10^{-3}$, $n < 4$, $h_n = 10^{-1}$, $n \geq 4$.

x	y_1	y_2	y_3
0.4	$.7 \cdot 10^{-7}$	$.1 \cdot 10^{-6}$	$.2 \cdot 10^{-6}$
10	$.7 \cdot 10^{-6}$	$.1 \cdot 10^{-7}$	$.7 \cdot 10^{-6}$

Stapkeuze strategie bij eferk

In de procedure eferk wordt een andere stapkeuze strategie gebruikt dan bij efsirk. Wij gebruiken in eferk namelijk een schatting van de locale fout, die in het geval van lineaire vergelijkingen slechts 4^e en hogere orde termen bevat, bij niet-lineaire vergelijkingen ook een 3^e orde term. Beschouw de residu functie (Van der Houwen, Beentjes, Dekker en Slagt (1971), pagina 60 en volgende)

$$(8.38) \quad \xi_n(h) = \frac{d}{dh} y_{n+1}(h) - f(y_{n+1}(h)),$$

waarin $y_{n+1}(h)$ de numerieke oplossing voorstelt, indien uitgaande van y_n geïntegreerd wordt met staplengte h . In bovengenoemde referentie is afgeleid dat $h_n \mid \xi_n(h_n) \mid$ een goede schatting van de lokale fout voorstelt. We definiëren nu

$$(8.39) \quad \rho_n^*(h) = h \mid R'(hJ_n) f(y_n) - f(y_{n+1}) \mid.$$

We zullen aantonen dat $\rho_n^*(h_n)$ een goede benadering is van de lokale fout. Voor autonome lineaire vergelijkingen geldt namelijk

$$\rho_n^*(h_n) = h_n \mid \xi_n(h_n) \mid;$$

Dit volgt onmiddellijk door differentiatie naar h van de gelijkheid

$$y_{n+1}(h) = y_n + \{R(hJ_n) - 1\}J_n^{-1} f(y_n)$$

en substitutie in (8.38). Op dezelfde wijze kan eenvoudig worden geverifieerd dat bij niet-lineaire vergelijkingen geldt

$$\rho_n^*(h_n) = h_n \mid \xi_n(h_n) \mid + O(h_n^3).$$

(vergelijk ook de opmerking).

Evenals bij de procedure efsirk kan nu uitgaande van h_n , tol en $\rho_n^*(h_n)$ een nieuwe stap h_{n+1} gekozen worden. In de procedure eferk luidt de uitdrukking voor h_{n+1}

$$(8.40) \quad \begin{cases} h_{n+1} = \left\{ \frac{4 \text{ tol}}{3 \text{ tol} + \rho_n^*(h_n)} + \frac{1}{3} \right\} h_n, & \text{bij niet-lineaire vergelijkingen} \\ h_{n+1} = \left\{ \frac{4 \text{ tol}}{4 \text{ tol} + \rho_n^*(h_n)} + \frac{1}{2} \right\} h_n, & \text{bij lineaire vergelijkingen.} \end{cases}$$

Opmerking:

We kunnen $\rho_n^*(h_n)$ ook opvatten als de discrepantie tussen y_{n+1} en een referen-

tie-oplossing $\tilde{y}_{n+1}$, die dan voldoet aan de formule

$$(8.41) \quad \tilde{y}_{n+1} = y_{n+1} + h_n f(y_{n+1}) - R'(h_n J_n) h f(y_n).$$

Het stabiliteitspolynoom $\tilde{R}(z)$ van deze formule luidt

$$(8.42) \quad \tilde{R}(z) = R(z) + z(R(z) - R'(z));$$

We kunnen hieruit afleiden dat $\tilde{R}(z)$ ook 3e orde consistent is, en in dezelfde punten als $R(z)$ een exponentiële aanpassing heeft.

Ontwikkelen we $\tilde{y}_{n+1}$ in een Taylor-reeks, dan vinden we

$$(8.43) \quad \tilde{y}_{n+1} = y_n + h_n f(y_n) + \frac{1}{2} h_n^2 J_n f(y_n) + \frac{1}{6} h_n^3 J_n^2 f(y_n) + \frac{2}{3} h_n^3 y_n''' + O(h_n^4).$$

Voor niet-lineaire vergelijkingen volgt hieruit

$$|\tilde{y}_{n+1} - y_{n+1}| = \frac{1}{2} h_n^3 y_n''' + O(h_n^4),$$

terwijl we voor lineaire vergelijkingen uit (8.42) afleiden

$$|\tilde{y}_{n+1} - y_{n+1}| = h_n^4 J_n^3 f(y_n) \left(\frac{1}{6} - \beta'_3 \right)$$

(Hierbij is β'_3 de coëfficiënt van z^3 uit $R'(z)$).

Met deze stapkeuze strategie is probleem (8.25) nogmaals opgelost. Voor de minimale staplengte kozen we $h_{\min} = .001$, voor de maximale staplengte $h_{\max} = \text{if } x = 0 \text{ then } .001 \text{ else if } x < 5 \text{ then } .1 \text{ else } .2$. In tabel 8.14 staan de absolute fout in de componenten en het aantal integratie stappen vermeld.

Tabel 8.14

Absolute fout en het aantal stappen voor probleem (8.25).

x	abs. ϵ = rel. ϵ = 10^{-6}				abs. ϵ = rel. ϵ = 10^{-4}			
	y_1	y_2	y_3	n	y_1	y_2	y_3	n
0.4	.3 10^{-7}	.1 10^{-7}	.2 10^{-7}	16	.2 10^{-6}	.1 10^{-6}	.1 10^{-6}	12
10	.2 10^{-6}	.1 10^{-7}	.2 10^{-6}	142	.2 10^{-5}	.1 10^{-7}	.2 10^{-5}	85

Verder hebben we als testvoorbeelden de differentiaal-vergelijkingen (8.33) en (8.34) beschouwd. De resultaten voor diverse waarden van de tolerantie staan in de tabellen 8.15 en 8.16 vermeld. (In beide gevallen kozen we $h_{\min} = .005$ en $h_{\max} = .5$).

Tabel 8.15

$-^{10}\log$ rel. fout en het aantal stappen voor probleem (8.33).

x	abs ϵ = rel ϵ = 10^{-2}	n	abs ϵ = rel ϵ = 10^{-5}	n
0.4	3.0	13	5.1	70
1	4.5	88	6.9	408

Tabel 8.16

$-^{10}\log$ rel.fout en het aantal stappen voor probleem (8.34).

x	abs ϵ = rel ϵ = 10^{-2}	n	abs. ϵ = rel ϵ = 10^{-5}	n
0.5	3.4	2	6.0	15
1	2.8	4	5.8	29

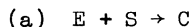
Literatuur

- Calahan, D.A. (1968), A stable, accurate method of numerical integration for non-linear systems, Proc. IEEE 56, 744.
- Fowler, M.E. and R.M. Warten (1967), A numerical integration technique for ordinary differential-equations with widely separated eigenvalues, IBM Journal, 537-543.
- Houwen, P.J. van der, P. Beentjes, K. Dekker and E. Slagt (1971), One step methods for linear initial value problems III, Numerical results, Report TW 130, Mathematisch Centrum, Amsterdam.
- Houwen, P.J. van der (1972a), Explicit and semi-implicit Runge-Kutta formulas for the integration of stiff equations, Report TW 132, Mathematisch Centrum, Amsterdam.
- Houwen, P.J. van der (1972b), Wordt gepubliceerd in de lecture notes van de conferentie over "Numerische, insbesondere Approximationstheoretische Behandlung von Funktionalgleichungen" te Oberwolfach, 1972.
- Lapidus, L. and J.H. Seinfeld (1971), Numerical solution of ordinary differential equations, Academic Press, New York.
- Liniger, W and R.A. Willoughby (1970), Efficient integration methods for stiff systems of ordinary differential equations, SIAM J., Numer. Anal. 7, 47.
- Robertson, H.H. (1967), The solution of a set of reaction rate equations in "Numerical Analysis", (J. Walsh, ed.), Thompson Book Co., Washington.
- Rosenbrock, H.H. (1963), Some general implicit processes for the numerical solution of differential equations, Comp. J. 5, 329.

9. Stijve differentiaalvergelijkingen in biochemische modellen

9.1. Inleiding

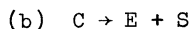
De biochemie zou een bevoorrechte discipline binnen de biologie genoemd kunnen worden. Op vele gebieden is het vaak lastig om aan te tonen dat een bepaald mathematisch model een zinvolle weergave is van een groep biologische fenomenen. De biochemie echter maakt gebruik van modellen die berusten op de aanname dat een chemische reactie plaatsvindt als gevolg van de botsing tussen twee molekulen. Deze modellen hebben een zeer grote geloofwaardigheid. Men kan namelijk stellen dat de reaktiesnelheid evenredig is met de botsingskans en daardoor met de concentraties van de betrokken stoffen. De chemische vergelijking



leidt daarom tot de formule voor de reaktiesnelheid

$$(9.1.1) \quad -\frac{dE}{dt} = -\frac{dS}{dt} = +\frac{dC}{dt} = k E S \quad . \quad *)$$

De omzettingssnelheid van één molekuulsoort is evenredig met zijn concentratie, dus als



dan geldt:

$$(9.1.2) \quad -\frac{dC}{dt} = +\frac{dE}{dt} = +\frac{dS}{dt} = lC \quad .$$

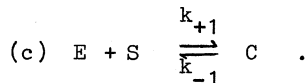
De evenredigheidsfactoren k en l (de reaktiekonstantes) zijn binnen één proef meestal onveranderlijk. Hun waarde wordt bepaald door die omstandigheden die bij vaste concentratie van de reaktanten maken dat

- a) er meer of minder botsingen zijn, en
- b) de kans dat een botsing succesrijk is.

Dat zijn dan bijvoorbeeld ad a) druk en temperatuur en ad b) temperatuur en thermodynamische eigenschappen van de reaktanten.

*) In chemische formules (aangegeven met letters) duiden E, S etc. een molekuul-soort aan, in de wiskundige formules de concentratie van die molekuulsoort.

Deze konstantes worden gewoonlijk weergegeven door k_j , waar j verwijst naar een bepaalde reactie. Over het algemeen zullen reacties reversibel zijn; dit kan men dan als volgt in een vergelijking schrijven

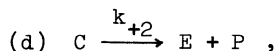


Als er evenwicht is geldt

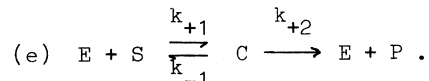
$$(9.1.3) \quad \frac{dC}{dt} = \frac{dE}{dt} = \frac{dS}{dt} = 0$$

zodat $k_{+1} ES = k_{-1} C$. Het quotiënt van de reactiekonstanten noemt men de evenwichtskonstante; deze noteert men met een hoofdletter K , dus bijvoorbeeld $K_1 = k_{-1}/k_{+1}$.

Als C nog tot een tweede reactie in staat is:



ontstaat het volgende schema:



Zelfs dit eenvoudige schema leidt al tot een analytisch niet oplosbaar, in het algemeen stijf, niet-lineair stelsel differentiaalvergelijkingen. Dit systeem is gelijk aan het systeem beschreven op blz. 11 e.v. van dit colloquiumverslag (E =enzym, S =substraat, C =enzymsubstraatcomplex, P =produkt). Een interessant aspect van een dergelijke reactievergelijking is dat k_{+1} en k_{-1} vaak veel groter zijn dan k_{+2} ; dit maakt dat gedurende een zekere tijd C praktisch even snel zal ontstaan als het wordt afgebroken omdat de konsentratie C blijft stijgen totdat de afbraaksnelheid (dat is $(k_{-1} + k_{+2})C$) gelijk is aan de aanmaaksnelheid $k_{+1}ES$. Men spreekt dan van een *steady state* van het systeem.

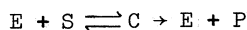
Aan de ene kant verkeert de biochemie in de gelukkige omstandigheid dat, op eenvoudige en eenduidige wijze, voor een fysisch-chemisch model een mathematische vergelijking kan worden opgesteld. Anderzijds zijn de

in de biochemie voorkomende katalysatoren - de enzymen - zeer complexe eiwitmolekullen die tot zeer ingewikkelde interacties in staat zijn. Bovendien werken dergelijke enzymen in de levende cel als onderdelen van grotere functionele eenheden, waarbij bijvoorbeeld het produkt van een enzym het substraat voor het volgende enzym is. Dit geeft aanleiding tot een grote klasse complexe wiskundige modellen die, voor een wiskundige, bijzonder interessante aspecten kunnen bezitten.

Het is de bedoeling van dit artikel wiskundigen een indruk te geven van het soort wiskundige problemen waarvoor de oplossing van belang is voor de biochemie, terwijl het anderzijds voor de biochemikus een exploratie en/of samenvatting kan zijn van de meer wiskundige aspecten van zijn vak.

9.2. Het bepalen van de reaktieconstanten van enzymatische reacties in het steady-state gebied

Op blz. 11 is reeds vermeld hoe men onder bepaalde voorwaarden het kinetische gedrag van het reactieschema



kan benaderen via de singuliere storingsrekening. Men vindt dan voor de reaktiesnelheid (bij gegeven E en S)

$$(9.2.1) \quad v = \frac{dP}{dt} = \frac{k_{+2}ES}{K_m + S}$$

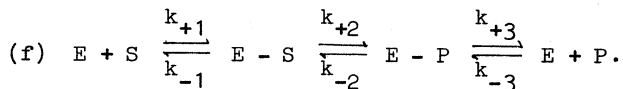
waarin $K_m = (k_{-1} + k_{+2})/k_{+1}$.

Tevens ziet men (blz. 15 vgl. (1.10)) dat de maximale waarde van v kort na het begin van de reactie bereikt wordt en dat v gedurende vrij lange tijd nagenoeg konstant blijft (N.B. de stijfheid van de differentiaalvergelijkingen die het systeem beschrijven is een begrijpelijk gevolg van dit gedrag).

De experimentele gegevens bestaan meestal uit een aantal metingen van v (de steady state snelheid) voor verschillende waarden van E en S.

Op grond van deze metingen kan men m.b.v. formule 9.2.1 slechts twee parameters schatten (K_m en k_{+2}), terwijl in het complete systeem drie parameters voorkomen. Wil men inzicht verkrijgen in het gebeuren op moleku-

lair niveau, dan is het van belang dat men alle drie de k 's kent. Het gebruikelijke soort biochemische metingen levert ons dus één parameter te weinig. De situatie is zelfs nóg ongunstiger: reaktieschema e) geeft veelal een onvoldoend realistische beschrijving van de verschijnselen. Men moet zich voorstellen dat het enzymsubstraatkomplex ontstaat, dat er zich vervolgens binnen dit complex verschuivingen van chemische bindingen voordoen waardoor het substraat in het produkt wordt omgezet en dat daarna het produkt uit het kompleks loslaat. Verder zijn het merendeel van de biochemische reakties reversibel, zodat de betiteling substraat of produkt slechts afhangt van de vraag welke stof er aan het begin van de proef in hoge konsentraties aanwezig is. Een meer realistisch reaktieschema dan e) ziet er daarom als volgt uit:



Als men, zoals wij voor reaktieschema e) gedaan hebben, aanneemt dat de komplexen in een steady state voorkomen, kan men ook hier benaderingsformules opstellen. Door afwisselend S en P als substraat te gebruiken kan men twee series reaktiesnelheden meten, nl. $v=f(s)$ en $g=g(p)$ en daaruit waarden vinden voor V_s , K_{ms} , V_p , en K_{mp} met behulp van de volgende uitdrukkingen

$$(9.2.2) \quad v_{(p=0)} = \frac{V_s \cdot S}{S + K_{ms}} \quad v_{(s=0)} = \frac{V_p \cdot P}{P + K_{mp}}$$

Het laat zich geheel analoog aan de ontwikkeling die tot (9.2.1) leidde, afleiden dat geldt:

$$(9.2.3) \quad \begin{aligned} V_s &= k_{+2}k_{+3} / (k_{+2} + k_{-2} + k_{+3}) E, \\ V_p &= k_{-2}k_{-1} / (k_{-2} + k_{+2} + k_{-1}) E, \\ K_{ms} &= \{k_{-1}(k_{-2} + k_{+3}) + k_{+2}k_{+3}\} / \{k_{+1}(k_{+2} + k_{-2} + k_{+3})\}, \\ K_{mp} &= K_{ms} (P V_p) / (S V_s). \end{aligned}$$

De steady state behandeling van dit probleem geeft slechts vier van de zes parameters die ons interesseren. Het feit dat wij hier met een benaderende formule goed toekomen wordt veroorzaakt doordat de eigenschap $k_{+1} \gg k_{+2}$ maakt dat, op de experimenteel toegankelijke tijdschaal, het systeem zich eenvoudiger gedraagt dan uit de algemene formule zou volgen. Men kan dan ook zonder speciale voorzorgen geen meetwaarden verzamelen die informatie geven over de parameters welke in de benaderende formules niet onafhankelijk voorkomen.

Voor de mathemतिकus is dit niet verbazingwekkend. Wat de biochemikus de steady state benadering noemt is het gebied waarin de oplossing van de gereduceerde vergelijking van het singuliere storingsprobleem een benadering levert die binnen de grenzen van de experimentele fout ligt. De experimentele fout ligt tussen 1% en 5% en is dus vrij groot. Hierdoor zal in een groot gebied deze eerste benadering voldoende zijn. De afwijkingen van deze benadering, die inlichtingen zouden kunnen geven over de ontbrekende parameters gaan in diezelfde experimentele fout onder.

Waarom willen wij echter alle reaktiekonstanten weten ?

Hiervoor zijn twee redenen:

1. Uit variatie van reaktiekonstanten met externe omstandigheden als temperatuur, druk, zuurgraad, etc. kunnen chemici een indruk krijgen hoe het precieze verloop is van de interactie tussen enzym en substraat.
2. De waarden voor k_{+2} en K_m geven alleen een voldoende beschrijving onder de kondities waaronder zij verkregen zijn; d.w.z. overmaat substraat, afwezigheid van produkt etc.. Deze omstandigheden zijn vaak niet de omstandigheden waaronder het enzym in de levende cel werkt, de omstandigheden dus, die voor de bioloog van belang zijn.

In de praktijk zijn ook bij de eenvoudigste enzymatische reacties meestal twee substraten betrokken. Het reaktieschema wordt daarmee uitgebreider. De experimentele mogelijkheden verruimen daarmee eveneens. Altijd echter ziet men dat men onder steady-state kondities voor iedere extra experimenteel te bepalen parameter twee extra theoretisch gefundeerde parameters moet invoeren (zie ook sub 3).

9.3. Waarnemingen buiten het steady-state gebied

Het meten van de individuele reaktiekonstanten is mogelijk door het verrichten van waarnemingen in de zogenaamde *transient state*, d.w.z. in het gebied waar de reaktiesnelheid bij gegeven beginwaarden van E en S nog niet zijn maximale waarde heeft bereikt. Aangezien de klassieke enzympkinetische steady-state benadering dan niet meer voldoende is, werkt men op een tijdschaal waar de reaktiesnelheid essentieel van alle reaktiekonstanten afhangt. De voor de hand liggende oplossing, het meten van v als functie van t, E en S, kan inderdaad goed dienen om alle reaktiekonstanten te meten (ref. 13), maar stuit op drie soorten moeilijkheden, nl.

a) experimentele, b) mathematische, en c) statistische.

ad a. De experimentele moeilijkheden:

De experimentele problemen vinden hun oorzaak in het feit dat de transient state meestal snel verloopt en vaak van de orde van grootte van millisekonden is. Het vereist heel aparte technieken om in dit gebied metingen te kunnen doen. Dit probleem heeft men op twee manieren trachten op te lossen.

- 1) Bij de zogenaamde stopped flow methoden mengt men, in een speciaal instrument, razendsnel de reagentia om daarna langs optische weg het reaktieverloop te volgen (ref. 6).
- 2) Bij de zogenaamde relaxatiemethoden (ref. 4) gaat men uit van een reaktiemengsel in evenwicht (b.v. het mengsel van reaktieschema (f) en brengt dit dan zeer snel (door een temperatuur- of drukverandering) uit zijn evenwicht. Uit de wijze waarop het systeem dan zijn nieuwe evenwicht bereikt, kan men de reaktiekonstanten berekenen (zie onder).

ad b. De mathematische problemen:

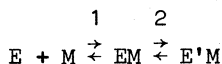
Een goede analyse van een stopped flow experiment vereist een oplossing van de niet-lineaire differentiaalvergelijkingen, die in het eenvoudigste geval volgen uit reaktieschema (e). Dit zijn, door het uiteenliggen van de verschillende tijdschalen, juist stijve differentiaalvergelijkingen. Uitgebreider modellen zijn nog moeilijker oplosbaar. Dit

probleem is door de biochemici die deze methoden gebruiken tot nu toe niet adequaat opgelost (ref. 5). De meestal inadequaatte aanpak van de vergelijkingen maakt veel van het tot nog toe in de biochemische literatuur op dit gebied verschenen mathematische werk diskutabel.

Bij de relaxatiemethoden is het essentieel dat het systeem slechts weinig uit zijn evenwicht wordt gebracht. Dit rechtvaardigt binnen zekere grenzen linearisatie van de differentiaalvergelijkingen, en omzeilt daarbij vele mathematische problemen. De prijs wordt betaald in de vorm van hoge eisen aan - en beperking van - de experimentele omstandigheden, nl.

- a) het systeem moet, voor de proef begint, in evenwicht zijn; reaktieschema's van het type e) kunnen dus niet worden bestudeerd, van het type f) wel;
- b) men moet een interne thermodynamische variabele (druk, temperatuur) kunnen variëren als een stapfunctie, in ieder geval op een tijdschaal welke een orde kleiner is dan de kleinste tijdschaal van het systeem. Dit vereist een bijzonder experimentele opzet;
- c) de kleine verschuiving uit evenwicht moet waarneembaar zijn.
Men moet daarom werken bij hoge enzymkonsentrasië en er moet überhaupt een fysische parameter (meestal een optische) zijn die met een variëren-de konsentratië in het reaktiemengsel mee varieert.

Hoe de analyse verloopt in een eenvoudig geval dat bovendien aan alle experimentele eisen voldoet zien we in het volgende voorbeeld (ref. 4).
Laten we de volgende reaktie beschouwen



en stel dat wij de konsentratië van de intermediären apart kunnen waarnemen. De snelheidsvergelijkingen zijn

$$(9.3.1) \quad \frac{dM}{dt} = -k_{+1} \cdot E \cdot M + k_{-1} EM$$

$$\frac{dEM}{dt} = k_{+1} \cdot E \cdot M - (k_{-1} + k_{+2}) EM + k_{-2} E'M$$

Als we de volgende notatie invoeren

$$\delta M = x_1, \quad \delta E = x_1,$$

$$\delta EM = x_2, \quad \delta E'M = -(x_1 + x_2),$$

$$\alpha_{11} = -k_{+1}(E+M), \quad \alpha_{12} = k_{-1},$$

$$\alpha_{21} = k_{+1}(E+M) - k_{-2}, \quad \alpha_{22} = -(k_{-1} + k_{+2} + k_{-2}),$$

zijn de gelineariseerde vergelijkingen

$$(9.3.2) \quad \frac{d}{dt} x_1 = \alpha_{11} x_1 + \alpha_{12} x_2$$

$$\frac{d}{dt} x_2 = \alpha_{21} x_1 + \alpha_{22} x_2.$$

Deze lost men op de gebruikelijke wijze op via de karakteristieke vergelijking

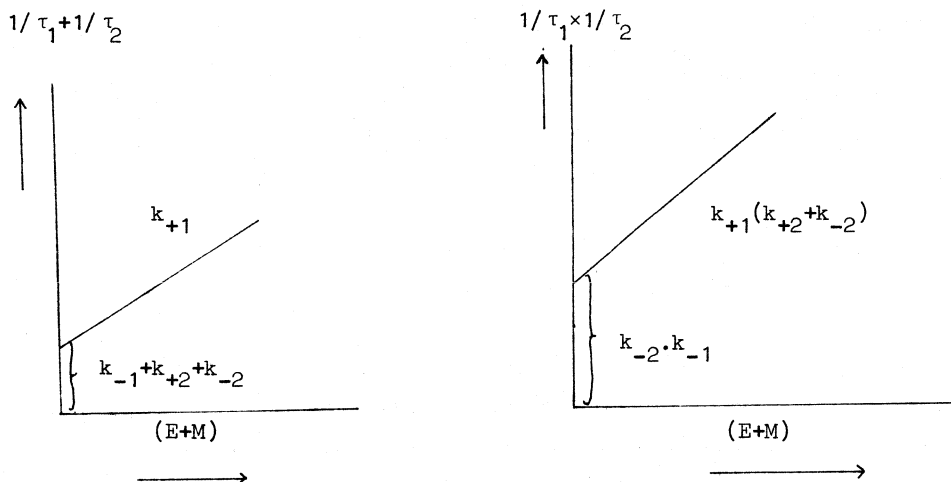
$$\begin{vmatrix} \alpha_{11} + 1/\tau_1 & \alpha_{12} \\ \alpha_{21} & \alpha_{22} + 1/\tau_2 \end{vmatrix} = 0.$$

Experimenteel neemt men waar hoe een konsentratie na verstoring zijn nieuwe waarde bereikt. Dit analyseert men als de som van een aantal (hier twee) exponentiële funkties. De tijdkonstanten van deze exponentiële funkties zijn de reciproke relaxatietijden, waarvoor geldt

$$(9.3.3) \quad \frac{1}{\tau_1} + \frac{1}{\tau_2} = -(\alpha_{11} + \alpha_{22}) = k_{+1}(E+M) + k_{-1} + k_{+2} + k_{-2},$$

$$\frac{1}{\tau_1} \times \frac{1}{\tau_2} = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} = k_{+1}(k_{+2} + k_{-2})(E+M) + k_{-1}k_{-2}.$$

Dit maakt het mogelijk aan de hand van de experimentele gegevens de volgende grafieken te konstrueren:



Uit dit soort grafiek leest men dan alle reaktieconstanten af. Vergelijkbare methoden kan men gebruiken voor veel ingewikkelder reaktieschema's. Deze (en dergelijke) grafieken leiden meteen het volgende probleem in: ad c. De statistische problemen:

Deze vinden hun oorzaak in het feit dat, ook al heeft men een theoretisch model en de komplette wiskundige beschrijving daarvan, de waarden van de parameters nog niet eenduidig uit de experimentele gegevens volgen. De experimentele gegevens zijn belast met meetfouten en men neemt veelal aan dat de meetfout een normale verdeling heeft. Door een serie experimentele punten moet nu de best passende curve getrokken worden. Deze curve is een oplossing van het stelsel differentiaalvergelijkingen dat kan worden opgesteld aan de hand van het model. Als de experimentele fout voor ieder punt bekend is, is het mogelijk te schatten hoe goed de aanpassing van de curve kan zijn. Als de werkelijk bereikte aanpassing hier significant van afwijkt is dit een aanwijzing dat het model niet juist is. Op deze wijze kan men uit de verwerking van experimentele gegevens aanwijzingen verkrijgen over de onjuistheid van een model. Het ligt in de aard van deze benadering dat men nooit de juistheid van een model kan bewijzen. Ten hoogste kan men bewijzen het eenvoudigste model gevonden te hebben dat bepaalde experimenten verklaart.

Samenvattend kan men dus als belangrijkste wiskundige probleem aanmerken het oplossen van de niet-lineaire stijve differentiaalvergelijkingen van het type dat ontstaat uit de reaktievergelijkingen zoals weergegeven in e) en f). De oplossing zou dan gebruikt kunnen worden voor de volgende deelproblemen:

- a. analyse van stopped flow experimenten;
- b. analyse van relaxatie-experimenten buiten het lineaire gebied.

Als men beschikt over een juiste wiskundige analyse zou men sommige experimentele voorwaarden bij relaxatie-experimenten kunnen laten vallen. Het is dan bijvoorbeeld niet meer nodig dat de verstoring klein is, omdat linearisatie van de differentiaalvergelijkingen geen voorwaarde meer is voor oplossing. Ook zou men kunnen nagaan hoe de methoden van de relaxatie-analyse zouden kunnen worden toegepast op een in gang zijnde reactie. Hiermee vervalt ook de noodzaak om in een relaxatiesysteem alleen evenwichtsreacties te beschouwen.

9.4. Ingewikkelder reaktieschema's met één enzym (in niet koöperatieve systemen)

Tot nu toe hebben wij slechts het eenvoudigste reaktieschema behandeld, dat voor één enzym mogelijk is. In de praktijk zijn de reaktiemechanismen veel ingewikkelder, omdat er meer dan één substraat kan optreden of omdat er stoffen kunnen zijn die een enzym remmen of stimuleren. Verder bestaat de mogelijkheid dat een overmaat van substraat (c.q. produkt) remt (c.q. aktiveert). Het aantal mogelijkheden is legio. De methode van oplossing is in principe gelijk aan die welke voor één substraat is aangegeven:

- a) men stelt reaktievergelijkingen op;
- b) van ieder van de componenten geeft men de konsentratieverandering per tijdseenheid aan door middel van een stelsel differentiaalvergelijkingen;
- c) men stelt de zogenaamde behoudsvergelijking op. Deze vergelijkingen ont-

staan doordat de som van de konsentratie van alle vormen van enzymcomplexen altijd gelijk is aan de oorspronkelijk aanwezige enzymkonsentratie. Op ieder ogenblik is de som van een substraatkonsentratie plus de konsentratie van het daaruit ontstane produkt plus alle complexen waaraan het deelneemt gelijk aan de beginkonsentratie van dat substraat etc.. Deze vergelijkingen geven een koppeling tussen de differentiaalvergelijkingen. De beginvoorwaarden zijn in het eenvoudigste geval

$$E=E_{\text{totaal}}, S=S_{\text{totaal}}, P=0, C=0 \text{ etc..}$$

Een eerste benaderende oplossing verkrijgt men (ook voor complexe systemen) altijd via een steady-state methode (d.i. de singuliere storingsrekening). Hiertoe stelt men voor ieder van de enzymsubstraatcomplexen de verandering in de tijd nul, en lost de resterende vergelijkingen op. Meestal nog onder de bijkomende aanname dat de substraatkonsentratie veel groter is dan de enzymkonsentratie. King and Altman (ref. 9) hebben het principe van een algemene methode aangegeven om de kinetiek van ingewikkelde reaktieschema's volgens de steady-state methode uit te werken.

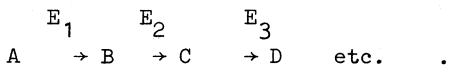
Er bestaat geen algemene methode voor het schrijven van de differentiaalvergelijkingen voor dergelijke methodes, en nog minder voor het vinden van hun analytische oplossing.

9.5. Enzymsystemen

Enzymen komen in de natuur in een funktioneel verband voor. Om een bepaald biologisch proces mogelijk te maken moet een enzym samenwerken met andere enzymen. Twee vormen van zo'n samenwerking zullen wij hier behandelen, de keten en de kaskade. Bij beide systemen gaan we ervan uit dat enzym, substraat, en produkt in een perfect geroerde oplossing aanwezig zijn, zodat de konsentraties niet plaats-afhankelijk zijn. In de praktijk is dit vaak niet het geval, daar spelen naast de puur chemische processen nog transportprocessen een rol. Niet verwaarlozen van de diffusie leidt tot aparte problemen, o.a. omdat nu de ruimtekooördinaten gaan meespelen. Deze zullen we hier niet behandelen (zie hiervoor bijvoorbeeld ref. 8).

9.5.1. Enzymketens

Men spreekt van een enzymketen als het produkt van één enzym het substraat is voor een volgend, schematisch



De reaktiesnelheid in een keten wordt vaak gereguleerd doordat een produkt (b.v. D) een voorgaand enzym (bijvoorbeeld E_1) remt. Dikwijls is hier sprake van allosterische remming of aktivatie (zie sub 6). Verder kunnen dergelijke ketens vertakt zijn. In een steady-state van een onvertakte reaktieketen zal de hoeveelheid die per tijdseenheid omgezet wordt in ieder van de stappen, gelijk zijn. De concentraties van de stoffen B, C, E etc. zullen die waarden bereiken welke hun omzettingssnelheden gelijk maken.

Als E_3 een zeer actief enzym is, zal de concentratie van C zeer laag worden. Het is voorstelbaar (en er zijn in de praktijk voorbeelden van bekend) dat C zo laag wordt, dat de steady-state behandeling voor de stap E_3 C $\rightarrow$ D niet meer verantwoord is. (Dit komt omdat $C \gg E_3$ een van de voorwaarden is voor het afleiden van formule (9.2.1)). Ook kunnen de concentraties zo in grootte verschillen dat sprake is van stijve differentiaalvergelijkingen waardoor numerieke oplossing met behulp van standaardmethoden praktisch onmogelijk wordt. (Voor de vaak amusante noodsprongen waartoe biochemici gedwongen waren zie men bijvoorbeeld ref. 5). Voor ingewikkelder systemen, waar alle concentraties in principe alle snelheden kunnen beïnvloeden (door remming en aktivatie etc.) is de flux (F_i) een funktie van de enzymconcentraties (E_i) $i=1, \dots, n$ de substraatkonzentraties (S_i) en de kinetische konstanten (k_{i1} , k_{i2} etc.) nodig om het gedrag van enzym E_i te beschrijven. Verder zal de snelheid veelal evenredig blijven met de enzymconcentratie, zodat

$$F_i = E_i \cdot f_i(k_{ij} (j=1, \dots, r), S_i (i=1, \dots, n))$$

We kunnen schrijven

$$\frac{dS_k}{dt} = \sum \lambda_{ki} F_i$$

waarin $\lambda_{ki}=0$ als S_k niet reageert in reactie i . Als S_k substraat of produkt is in reactie i , dan is $\lambda_{ki}=-1$ respektievelijk $+1$, en als S_k remmer of aktivator is, geldt $\lambda_{ki} \approx 0$.

f_i is van de vorm van formule (9.2.2) of van de meer ingewikkelde vorm behandeld sub 4 of sub 6.

Het nut van numerieke simulatie ligt vooral hierin dat men een indruk kan krijgen van de "gevoelige plekken" in een enzym-netwerk, dat wil zeggen van die plaatsen waar een kleine verandering in enzymkonsentrasië of parameter een grote invloed heeft op het gedrag van het systeem (ref. 2).

In een enzymketen waar de eerste stap snelheidsbeperkend is of, algemener, in een keten achter een snelheidsbeperkende stap, geldt dat de substraatkonsentrasië vrij laag zal zijn. Toch is het mogelijk dat zij aanmerkelijk groter blijft dan E_i . Wij zijn dan in het gebied waar de reactiesnelheid evenredig is met de substraatkonsentrasië.

Als wij de evenredigheidskonstante aangeven met $1/\tau$, vindt men voor de konsentratiëverandering van het i -de substraat

$$(9.5.1) \quad \frac{dS_i}{dt} = \frac{S_{i-1}}{\tau_{i-1}} - \frac{S_i}{\tau_i} .$$

Dit heeft als algemene oplossing (ref. 3)

$$(9.5.2) \quad S_i(t) = v_0 \tau_i \left(1 - \sum_{j=1}^i A_{ij} e^{-t/\tau_j} \right) ,$$

v_0 is de snelheid van de eerste stap,

$$A_{ij} = \tau_j^{i-2} \prod_{\substack{k=1 \\ k \neq j}}^i \frac{1}{\tau_j - \tau_k} .$$

De steady state konsentratië wordt gegeven door

$$S_i = v_0 \tau_i .$$

De konsentratië van het produkt van een keten afgebroken op plaats i wordt gegeven door

$$(9.5.3) \quad p(t) = v_0 \left(t + \sum_{i=1}^n C_i e^{-t/\tau_i} - \sum_{i=1}^n \tau_i \right)$$

$$C_i = \tau_i^n \prod_{\substack{j=1 \\ j \neq i}}^n \frac{1}{\tau_i - \tau_j}$$

hetgeen voor $t \rightarrow \infty$ wordt

$$P = v_0 \left(t - \sum_{i=1}^n \tau_i \right) .$$

De algemene oplossing van het niet-vereenvoudigde geval

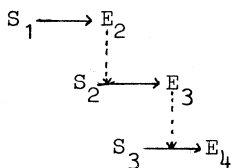
$$\frac{dS_i}{dt} = \frac{V_{i-1} \cdot S_{i-1}}{K_{i-1} + S_{i-1}} - \frac{V_i \cdot S_i}{K_i + S_i}$$

is nog niet bekend.

9.5.2 Enzymkaskades

Een enzymkaskade is een reeks enzymreacties waarin het produkt van een reactie het enzym (en niet het substraat) is voor de volgende. Men onderscheidt twee types, de open en de gedempte kaskade.

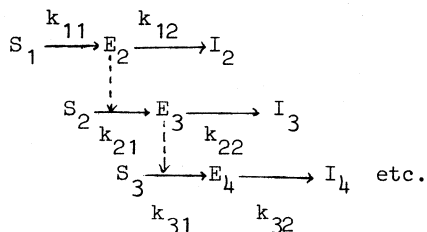
a) de open kaskade:



hier blijven alle enzymen behouden.

Vaak zien wij echter dat de enzymen weer onwerkzaam worden, bijvoorbeeld door de aanwezigheid van anti-enzymen, dit geeft de gedempte kaskade.

b. de gedempte kaskade.



In de open kaskade geldt voor de produktvorming

$$P(t) = (1/n)t^n E_1 \prod_{i=1}^n k_i$$

als wij aannemen dat $v = k E$ in iedere stap. Voor $v = \frac{kES}{K+S}$ wordt de formule:

$$P(t) = E_1 \prod_{j=1}^n \frac{S_j^{k+1} t^j}{S_j^{k+1} + K_{mj}} \times$$

(9.5.4)

$$\times \sum_{m=1}^n \left\{ \left(-\frac{1}{b_j} \right)^{n-m} A_j \right\} \frac{t^m}{m!} + \sum_{j=1}^n \left(-\frac{1}{b_j} \right)^n A_j (1 - e^{-b_j t}) ,$$

waarin

$$b_j = k + 1_j (S_j + k_{mj})$$

en

$$A_m = \prod_{\substack{j=1 \\ j \neq m}}^n \frac{b_j}{b_j - b_m} \quad (\text{N.B. } \sum_{m=1}^n A_m = 1) .$$

Voor de gedempte kaskade geldt, als $v = kE$

$$(9.5.5) \quad P(t) = E_1 \prod_{j=1}^n \frac{S_j^{k+1} t^j}{k + 2j} \sum_{j=1}^n A_j (1 - e^{-k+2j} t) ,$$

$$\text{waarin } A_j = \prod_{\substack{p=1 \\ j \neq p}}^n \frac{k+2j}{k+2j - k+2p} .$$

In het geval dat $v = \frac{kES}{K+S}$ geldt evenzo voor $P_n(t)$

$$P_n(t) = E_1 \prod_{j=1}^n S_j^{k+1} t^{k+2j} \left[\sum_{j=1}^n N_j (1 - e^{-k+3j} t) + M_j (1 - e^{-b_j} t) \right]$$

waarin

$$b_j = S + 1_j^{k+1} j^{k-1} j^{k+3j} ,$$

$$(9.5.6) \quad N_j = \frac{1}{k_{+3j}} \prod_{i=1 \atop i \neq j}^n \frac{1}{b_i - b_j} \prod_{j=1}^n \frac{1}{b_j + k_{+2j}}$$

$$M_j = \frac{1}{k_{-1j} + k_{+2j}} \prod_{i=1 \atop i \neq j}^n \frac{1}{b_i - b_j} \prod_{j=1}^n \frac{1}{k_{+3i} - b_j} . \quad (\text{ref. 7}) .$$

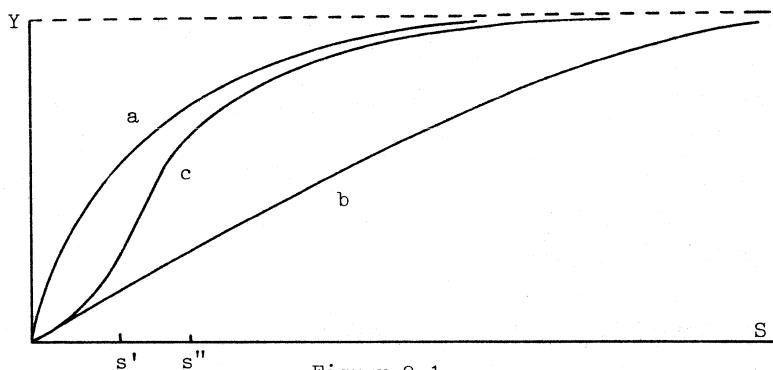
Kaskade systemen komen onder andere voor in de bloedstolling, bij immunologische systemen en misschien in de retina.

9.6. Koöperatieve systemen

Tot nu toe hebben wij reacties beschouwd waarbij de enzymmolekulen elkaar niet konden beïnvloeden. Er is echter een belangrijke groep biochemische reacties waarbij dit wel het geval is. Dit ziet men bij de zogenaamde *allostere enzymen*. Het is mogelijk dat deze enzymen voorkomen in de vorm van zogenaamde oligomeren. Deze bestaan uit een - meestal gering - aantal identieke enzymmolekulen (de subunits of protomeren) die hecht aan elkaar verbonden zijn en opereren als een eenheid. Als voorbeeld zullen wij behandelen een complex molekuul dat uit vier identieke onderdelen bestaat. Ieder van de protomeren heeft één bindingsplaats voor het substraat. De vier protomeren kunnen invloed op elkaar uitoefenen. Hierdoor zijn de eigenschappen van een protomeer dat niet met substraat is bezet verschillend naar gelang er geen, één, twee, of drie naastgelegen protomeren bezet zijn.

Wij zagen dat bij een enkelvoudig enzym de reactiesnelheid evenredig is met de concentratie van het enzym-substraatcomplex (C in reactievergelijking e); dat wil zeggen met de fraktie (Y) van de bindingsplaatsen die bezet is. Bij de allosterische enzymen is dat net zo. De hoeveelheid bindingsplaatsen die bezet is bepaalt de reactiesnelheid (naast natuurlijk de reactieconstante zoals k_{+2} in de reactie $C \xrightarrow{k_{+2}} E + P$). Wij hebben gezien dat de fraktie van de bindingsplaatsen die bezet is, een gebroken lineaire funktie is van de substraatkonsentratie. Voor deze fraktie Y geldt bij een enkelvoudig enzym

$$Y = \frac{C}{E} = \frac{V}{V_{\max}} = \frac{S}{K+S} .$$



Figuur 9.1

In figuur 9.1 geeft curve a het gedrag van een enkelvoudig enzym met kleine K en curve b dat van een enzym met grote K. Het interessante van allosterische enzymen is nu, dat K verandert met het aantal plaatsen dat bezet is. Als een volkomen "leeg" enzym met bijvoorbeeld vier bindingsplaatsen de grote K heeft van curve b in figuur 9.1, terwijl het nagenoeg geheel bezette enzym de kleine K heeft van curve a, dan zal de bezettingsgraad van het allosterische enzym door een curve als c weergegeven worden.

Op het ogenblik zijn twee modellen in omloop voor de verklaring van dit gedrag, het model gelanceerd door Adair et al. (ref. 1,10), dat wij het A-model zullen noemen, en dat van Monod et al. (ref. 11), het M-model.

Het A-model gaat ervan uit dat door de binding van één substraat-molekuul de bindingskonstanten van de overige vrije substraatbindingsplaatsen op het enzym veranderen.

Het M-model berust op drie aannamen:

- a. Het enzym bestaat uit subunits, die ieder het substraat kunnen binden;
- b. Het enzym kan in twee toestanden voorkomen (R en T) die in evenwicht

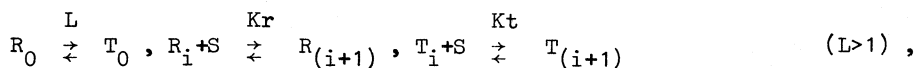
zijn, maar die verschillen in hun affiniteit tot het substraat;

c. Alle subunits van een molecuul zijn tegelijk in de T of in de R toestand.

Zonder substraat is het enzym bij voorkeur in de T-toestand ($L > 1$), maar het substraat bindt liever aan de R vorm en houdt het enzym in die vorm ($K_r > K_t$).

Hierdoor fixeert de eerste kleine hoeveelheid substraat de enzymen in hun "reageerbare" toestand.

In de vergelijkingen



zijn K_r en K_t ($K_r \neq K_t$) dus onafhankelijk van i . ($i=0, 1, \dots, n-1$; n = aantal subunits van het enzym).

Het zal duidelijk zijn dat beide modellen ("A" en "M") een sigmoïde kurve kunnen verklaren.

Voor de steady state laat zich nu de reaktiesnelheid uitrekenen

$$v = Y E_{\text{tot}} k_{+2},$$

waarin:

$$(9.6.1) \quad Y = \left(\sum_{i=0}^n i R_i + \sum_{i=0}^n i T_i \right) / n \cdot E_{\text{tot}}.$$

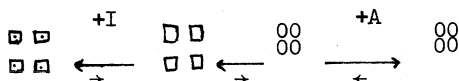
Wij rekenen verder met deze Y , omdat E_{tot} en k_{+2} hier geen ter zake doende parameters zijn.

Voor Y geldt:

$$(9.6.2) \quad Y = \frac{L\alpha(1+\alpha)^{n-1} + \alpha(1+\alpha)^{n-1}}{L(1+\alpha)^{n+1} + \alpha^n}.$$

De zaak begint iets gekompliceerder te worden als zich remmers en aktivatoren gaan voordoen. Een allosterische remmer (I) of aktivator (A) moet

men zien als een klein-moleculaire stof die een enzym-monomeer tijdelijk in de T- of de R-vorm "bevroren", dit maakt het dus mogelijk die vorm te bevoordelen. Onder invloed van een aktivator verschuift daarom het evenwicht tussen T en R naar de R-zijde



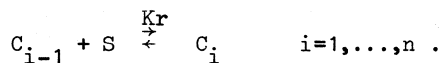
Ook hiervoor laat zich geheel analoog aan het bovenstaande een formule ontwikkelen; net zoals zonder modifikatoren zal gelden voor $c \rightarrow 0$

$$(9.6.3) \quad Y_s = \frac{\alpha(1+\alpha)^{n-1}}{L' + (1+\alpha)^n}.$$

Hier bepaalt L' het nieuwe evenwicht dat ontstaat onder invloed van de modifikatoren. Men kan eenvoudig laten zien dat

$$\begin{aligned}
 L' &= L \frac{(1+\beta)^n}{(1+\gamma)^n} \\
 (9.6.4) \quad &\text{met } \beta = \frac{I}{K_I} \quad \text{en} \quad \gamma = \frac{A}{K_a}.
 \end{aligned}$$

Het A-model lijkt eenvoudiger, de reactievergelijkingen zijn:



$$(9.6.5) \quad C_i = (S/K_i) C_{i-1}, \quad c_0 = E_{\text{vrij}},$$

$$C_i = S^i \prod_{j=0}^i 1/K_j E_{\text{vrij}},$$

$$Y = \sum_{i=1}^n i C_i / n E_{\text{tot}} \quad ,$$

(9.6.5)

$$Y = \sum_{i=1}^n \frac{i S^i}{n E_{\text{tot}}} \prod_{j=0}^n \frac{1}{K_j} .$$

K_i is de evenwichtskonstante voor de i -de reactie. Een positief koöperatief effect wordt verkregen als K_i daalt met stijgende i . Negatieve koöperatie, die niet met het M-model kan worden verklaard, ontstaat als K_i stijgt als i stijgt. Verdere precisering van de formules is moeilijk als men niet weet hoe de functie $K_i = f(i)$ verloopt.

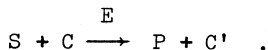
Het is duidelijk dat zowel met het M- als met A-model $v = f(s)$ een polynoom kan worden.

De relaxatiemethoden van Eigen kunnen ook op deze systemen worden toegepast. Zie hiervoor verder ref. 4.

Meer algemene oplossingen dan de steady-state vergelijkingen en de gelineariseerde relaxatievergelijkingen zijn niet bekend.

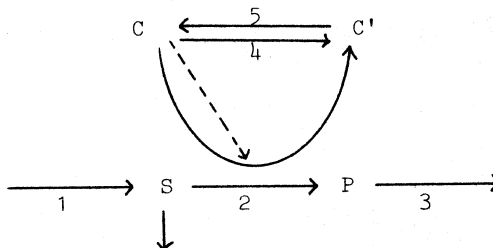
9.7. Niet-lineaire multi-enzym systemen

Uit de vorige paragraaf blijkt dat $\frac{dS}{dt}$, als functie van S , in teller en noemer niet-lineaire termen kan bevatten t.g.v. allosterie. Ook andere mechanismen kunnen niet-lineariteit veroorzaken. Wij hebben al aangestipt dat vele enzymatische reacties twee substraten hebben. Dit kan beschreven worden met de volgende vergelijking



C is vaak een "co-enzym", een organische stof die onder invloed van vrij veel verschillende biochemische rakties kan worden omgezet. De omzettingen $C \longrightarrow C'$ en $C' \longrightarrow C$ kunnen door verschillende enzymen worden verricht (v_4 en v_5 , zie onder). Daarnaast kan de omzetting $S \rightarrow P$ een onderdeel van een enzymketen zijn. Nog een derde mechanisme kan niet-lineariteit veroorzaken. Het enzym E kan geremd worden door een overmaat C . Dit geeft diagrammatisch het volgende reactieschema:

(9.7.1)



(De stippellijn duidt op remming).

Men kan aantonen dat in een dergelijk systeem (ref. 12) de reaktiesnelheid als het volgt benaderd kan worden:

$$(9.7.2) \quad v_2 = \frac{V_{SC}}{K_1(K_2 + C(1 + (\frac{C}{K_i})^n))}$$

waarin $V = k_{+2}E$; K_1 is de Michaelis-konstante voor S, K_2 die voor C, K_i is de remmingskonstante, dat wil zeggen een konstante die de effectiviteit van C als remmer van zijn eigen omzetting weergeeft. De orde van de allosterische interactie wordt aangegeven door n ($n=2,3,4,\dots$). Deze formulering geeft een benadering voor de snelheid van reactie $S \rightarrow P$.

Betere (en daarmee meer ingewikkelde) benaderingen kan men afleiden aan de hand van de vorige paragraaf.

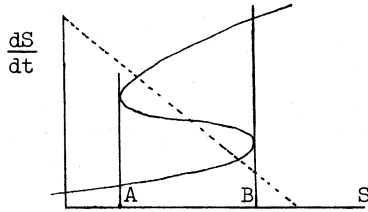
De overige vergelijkingen zijn

$$v_1 = V_m - K_1 S,$$

$$v_5 = K_5 C' \quad \text{en} \quad v_4 = K_4 C.$$

Verder geldt dat $C' + C$ konstant is. We houden de mogelijkheid open dat S via een andere reactie verdwijnt evenredig met zijn concentratie. Men kan dan een stelsel differentiaalvergelijkingen opstellen, dat het gedrag van dit systeem weergeeft. Dit gedrag blijkt merkwaardige eigenschappen te hebben (zie voor details ref. 12).

a. De $S-\frac{dS}{dt}$ karakteristiek kan sigmoïde zijn (zie fig. 9.3.).



Figuur 9.3. De $S-\frac{dS}{dt}$ karakteristiek voor systeem (9.7.1).

Hieruit blijkt dat het systeem een geheugen heeft: voor $A < S < B$ hangt de waarde van dS/dt van de voorgeschiedenis af. Als S eerst kleiner dan A was, dan is dS/dt klein, als S groter dan B was, zal dS/dt groot zijn.

b. Er zijn verschillende stationaire toestanden mogelijk, waaronder stabiele en niet-stabiele. Als de vormingssnelheid van S als functie van S wordt gegeven door de stippellijn in fig. 9.3, geven de snijpunten de stationaire substraatconcentraties weer. De twee buitensten daarvan zijn stabiel.

c. Een verstoring van een van de parameters die een bepaalde grens overschrijdt kan tot gevolg hebben dat dS/dt zeer sterk verandert, dit is mede afhankelijk van de duur van de verstoring. Kortdurende of kleinere verstoringen veroorzaken slechts een geringe afwijking van dS/dt . Dit gedrag is van groot belang in de electrofysiologie, b.v. voor de verklaring van het fysiologische fenomeen "prikkel drempel".

d. Het systeem kan een al dan niet stabiel "limit cycle" gedrag vertonen, met andere woorden, in het $S-\frac{dS}{dt}$ -diagram wordt een gesloten kurve met een bepaalde snelheid doorlopen. Van groot belang is, dat S nog aan een ander systeem van veranderingen kan worden onderworpen, nl. $S \xrightarrow{E'} S_d$ onder invloed van één enzym E' , en $S_d \xrightarrow{E''} S$ onder invloed van een ander enzym (E'').

Het stelsel differentiaalvergelijkingen zonder deze extra toevoeging luidt

$$(9.7.3) \quad \frac{dS}{dt} = v_1 - v_2$$

$$\frac{dC}{dt} = \frac{1}{\mu} (v_5 - v_4 - v_2)$$

waarin $v_2 = v_2(C)$.

Het is duidelijk dat afhankelijk van de waarde van μ deze vergelijkingen een stijf karakter kunnen hebben.

In de praktijk zien wij vaak dat de snelheid van de omzetting $S \rightarrow S_d$ en $S_d \rightarrow S$ veel langzamer is dan de snelheid van de andere vergelijkingen. Hierdoor verloopt deze omzetting weer in een geheel andere tijdschaal dan de twee bovengenoemde en het stijve karakter van het stelsel wordt nogmaals benadrukt.

In de laatste paragraaf is geen poging gedaan tot een vorm van analyse. De geïnteresseerde lezer zij verwezen naar ref. 12 voor een verdergaande bespreking. Het was de bedoeling erop te wijzen dat een iets ingewikkelder reaktieschema - dat overigens biochemisch zeer reëel is - behandeld volgens de principes uiteengezet in de voorgaande paragrafen, tot zeer interessante problemen kan leiden.

Literatuur

1. Adair G.S. (1925) J.Biol.Chem. 63, 529.
2. Burns J.A. (1969) FEBS Letters 2, 30.
3. Easterby J.S. (1973) Biochim. Biophys. Acta 293, 552.
4. Eigen M. (1968) Quart. Rev. Bioph. 1, nr.1, 3.
5. Garfinkel D., Garfinkel L., Pring M., Green S.B., Change B.
(1970) Ann. Rev. Bioch. 39, 473.
6. Gutfreund H. (1965) "An introduction to the study of enzymes",
Oxford, Blackwell Scient. Publ..
7. Hemker H.C., Hemker P.W. (1969) Proc. Roy. Soc. B. 173, 411.
8. Hermens W.Th., Hemker P.W., Hemker H.C. (1972)
in: "Analysis and Simulation of Biochemical Systems" p.19,
Amsterdam, North Holland Publ. Comp..
9. King E.L., Altman C. (1956) J. Phys. Chem. 60, 1375.
10. Koshland D.E., Nemethy G., Filmer D. (1966) Biochemistry 5, 365.
11. Monod J., Wyman J., Changeux J.P. (1965) J. Mol. Biol. 12, 88.
12. Selkov E.E. (1972) in "Analysis and Simulation of Biochemical
systems" p.145. Amsterdam, North Holland Publ. Comp.
13. Willems G.M., Hemker H.C. (1972) FEBS Letters 24, nr.3, 293.