MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
REKENAFDELING

CURSUS WETENSCHAPPELIJK REKENAAR(STER)

Numerieke Wiskunde

Deel II

door

J.A. Zonneveld

1957/1958

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

CURSUS WETENSCHAPPELIJK REKENAAR(STER)

Numerieke Wiskunde Deel II

door

J.A. Zonneveld

6. Approximatie

6.1 Inleiding

In dit hoofstuk zullen enkele methoden van approximatie van functies door andere functies behandeld worden.
Dit is belangrijk in het geval men een automatische rekenmachine dikwijls een ingewikkelde functie moet berekenen: men maakt dan een eenvoudig te berekenen approximatie. Naast deze approximaties komt veel voor dat een serie meetresultaten een functie van een bepaald type moet voorstellen. Het gaat er dan om de in die functie voorkomende parameters zo te bepalen, dat de aanpassing van functies aan meetresultaten in enig nader aan te geven zin, zo goed mogelijk is.

6.2. Methode van kleinste kwadraten

Bij deze methode wordt de som (of integraal) van het kwadraat van de afwijkingen geminimaliseerd.
Stel wij hebben m functiewaarden $f(x_i)$ $i = 1 \dots m$ en willen approximeren met een polynoom

$$\sum_0^n a_k x^k .$$

In een punt x_i is de afwijking $f(x_i) - \sum_0^n a_k x_i{}^k$. De som van de kwadraten van de afwijkingen, is:

$$F = \sum_{i=1}^{m} \left\{ f(x_i) - \sum_0^n a_k x_i{}^k \right\}^2 .$$

Opdat F minimaal zij, moeten tenminste alle differentiaalquotienten $\frac{\partial F}{\partial a_j} = 0$ zijn. Dat is op zichzelf niet voldoende voor een minimum. Het kan echter bewezen worden dat in dit geval aan alle verdere eisen voldaan is als geldt:

$$\frac{\partial F}{\partial a_j} = -2 \sum_{i=1}^{m} \left\{ f(x_i) - \sum_0^n a_k x_i{}^k \right\} . x_i^j = 0; \; j = 0(1)n \tag{6.2.1}$$

Hieruit volgen voor a_j de vergelijkingen:

$$\sum_{k=0}^{n} \left\{ \sum_{i=1}^{m} x_i^{\,k+j} \right\} a_k = \sum_{i=1}^{m} x_i^{\,j} f(x_i). \qquad j = 0(1)n \qquad (6.2.2)$$

of uitgeschreven:

$$(\sum_1^m x_i^0)a_0 + (\sum_1^m x_i^1)a_1 + \dots + (\sum_1^m x_i^n)a_n = \sum_1^m x_i^0\, f(x_i)$$
$$(\sum_1^m x_i^1)a_0 + (\sum_1^m x_i^2)a_1 + \dots + (\sum_1^m x_i^{n+1})a_n = \sum_1^m x_i^1\, f(x_i)$$

....

....

....

$$(\sum_1^m x_i^n)a_0 + (\sum_1^m x_i^{n+1})a_1 + \dots + (\sum_1^m x_i^{2n})a_n = \sum_1^m x_i^n\, f(x_i)$$

Nu kan wegens de bijzondere vorm van de coëfficiënten-matrix bewezen worden dat de determinant $\neq 0$ is: het stelsel vergelijkingen kan dus altijd opgelost worden, in theorie althans, want in de praktijk kan de oplossing zeer grote moeilijkheden opleveren. Hierover later meer.

Heeft men (6.2.2) opgelost dan heeft men dus het polynoom $\sum_0^n a_k x^k$ dat in de zin der kleinste kwadraten de gegeven rij functiewaarden $f(x_i)$ approximeert.

Bij het voorgaande is verondersteld dat het aantal gegeven punten, m groter is dan het aantal parameters, n+1. Is immers $m \leq n + 1$ dan kunnen wij de afwijkingen alle nul maken: wij hoeven dan slechts het interpolerend polynoom te bepalen:
Is $m < n + 1$ dan kunnen wij zelfs met een polynoom van de graad $m - 1$ volstaan.

voorbeeld

x	y	$p_2(x)$
-3	- .71	- .712
-2	- .01	- .002
-1	+ .51	+ .504
0	+ .82	+ .806
1	+ .88	+ .904
2	+ .81	+ .798
3	+ .49	+ .488

Gevraagd y te approximeren met een polynoom van de graad 2.

Wij vinden

$$\sum x_i^0 = 7;$$
$$\sum x_i^1 = 0;$$
$$\sum x_i^2 = 28;$$
$$\sum x_i^3 = 0;$$
$$\sum x_i^4 = 196;$$
$$\sum x_i^0 \; f(x_i) = 2.79;$$
$$\sum x_i^1 \; f(x_i) = 5.61$$
$$\sum x_i^2 \; f(x_i) = 2.61.$$

Het stelsel vergelijkingen wordt dan:

$$7a_0 + \quad + 28a_2 = 2.79$$
$$28a_1 = 5.61$$
$$28a_0 + 196a_2 = 2.61$$

en

$$a_0 = - .102$$
$$a_1 = + .200$$
$$a_2 = +.806$$

Hiermede volgen de opgegeven waarden $p_2(x) = a_0 + a_2x + a_2x^2$.

6.3 Bij approximatie van een functie (i.p.v. een rij functiewaarden) over een interval a - b, stellen wij vanzelfsprekend de eis dat

$$\int_a^b \left\{ f(x) - \sum a_k x^k \right\}^2 dx \text{ minimaal is} \qquad (6.3.1)$$

dan vinden wij i.p.v. de betrekkingen (6.2.1):

$$\int_a^b \left\{ f(x) - \sum a_k x^k \right\} x^j = 0; \; j = 0(1)n. \qquad (6.3.2)$$

Voor de verdere behandeling is het gemakkelijker om i.p.v. voor het interval a - b voor het interval 0 - 1 de formules af te leiden: door $t = \frac{x - a}{b - a}$ te stellen bereikt men dit. In het vervolg zal ons integratieinterval altijd 0 - 1 zijn.

Om de oplossing van het stelsel (6.3.2) te vergemakkelijken gaan wij heel anders te werk. Er worden ingevoerd polynomen $P_m(x)$, graad m zodanig dat $P_m(0) = 1$ is en verder geldt:

$$\int_0^1 P_m(x) P_n(x) dx = \begin{cases} 0 & \text{als } m \neq n \\ \frac{1}{2m + 1} & \text{als } m = n \end{cases} \qquad (6.3.3)$$

Wij stellen $P_m(x) = \sum_{k=0}^{m} a_k^{(m)} x^k \qquad (6.3.4)$

waarbij $a_0^{(m)} = 1$. Omgekeerd kan men iedere macht van x uitdrukken in een lineaire combinatie van $P_m(x)$. Onze approximatie $\sum_0^n a_k x^k$ kunnen wij dus ook schrijven als $\sum_0^n b_k P_k(x)$.

Doen wij dit, dan krijgen wij in plaats van (6.3.2)

$$\int_0^1 \left\{ f(x) - \sum_0^n b_k P_k(x) \right\} P_j(x) = 0, \; j = 0(1)n \qquad (6.3.5)$$

Met (6.3.3) volgt hieruit onmiddellijk:

$$b_j = (2j + 1) \int_0^1 f(x) P_j(x) dx. \qquad (6.3.6)$$

Door het invoeren van de functies $P_m(x)$ is het oplossen van het stelsel lineaire vergelijkingen een zeer eenvoudige zaak geworden! Alleen moet men achteraf nog de coëfficiënten a_k als lineaire combinatie van b_k's bepalen.

6.4. Polynomen van Legendre

Wij voeren in polynomen $P_m(x)$ van de graad m die voldoen aan

$$\int_0^1 P_m(x)x^s dx = 0 \text{ voor } s = 0(1)m - 1. \qquad (6.4.1)$$

met $P_m(0) = 1$.

Als $P_m(x) = \sum_{k=0}^{m} a_k^{(m)} x^k$ dan moet gelden volgens (1.4.1):

$$\sum_{k=0}^{m} \frac{a_k^{(m)}}{s + k + 1} = 0 \qquad \text{voor } s = 0(1)m - 1 \qquad (6.4.2)$$

Brengen wij alle termen van het linkerlid van (6.4.2) onder één noemer dan vinden wij voor dit linkerlid

$$\frac{a_0^{(m)}}{s + 1} + \frac{a_1^{(m)}}{s + 2} + \dots \frac{a_m^{(m)}}{s + m + 1} = \frac{Q_m(s)}{(s + 1)(s + 2) \dots (s + m + 1)} \qquad (6.4.3)$$

waarbij $Q_m(s)$ een polynoom in s van de graad m is. Volgens (6.4.2) heeft $Q_m(s)$ de nulpunten $s = 0(1)m - 1$ dus kan men schrijven met constante c

$$Q_m(s) = c\, s(s - 1)(s - 2) \dots (s - m + 1) \qquad (6.4.4)$$

of

$$c\,\frac{s(s - 1)\dots\dots\dots\dots(s - m + 1)}{(s + 1)(s + 2) \dots (s + m + 1)} = \frac{a_0^{(m)}}{s + 1} + \frac{a_1^{(m)}}{s + 2} + \dots + \frac{a_m^{(m)}}{s + m + 1} \qquad (6.4.5)$$

Vermenigvuldig beide leden van (6.4.5) met s + 1 en stel daarna s = - 1. Men vindt:

$$c\,\frac{-1.\ -2.\ \dots\ -m}{1.\ 2.\ \dots\ m} = a_0^{(m)} = 1$$

Dus $c = (-)^m$.

Door (6.4.5) aan beide zijden te vermenigvuldigen met s + k + 1 en daarna s = - k - 1 te stellen vindt men:

$$(-)^m \frac{- k - 1(- k - 2) \dots (- k - m)}{- k \dots - 1 \ 1 \dots (m - k)} = a_k^{(m)}$$

of $\quad a_k^{(m)} = (-)^k \binom{m + k}{k}\binom{m}{k} \qquad (6.4.6)$

Dus $P_m(x) = \sum_{k=0}^{m} (-)^k \binom{m+k}{k}\binom{m}{k} x^k$ (6.4.7)

Nu geldt volgens (6.4.1)

$$\int_0^1 P_m(x)P_m(x)dx = 0 \text{ als } m \neq n \qquad (6.4.8)$$

Als m = n dan weer volgens (6.4.1)

$$\int_0^1 P_m^2(x)dx = a_m^{(m)} \int_0^1 P_m(x)x^m dx$$

voor de tweede integraal vindt men door in het linkerlid van (6.4.5) s = m te stellen:

$$a_m^{(m)} \cdot (-)^m \cdot \frac{m!}{(m+1)(m+2)\ldots\ldots(2m+1)} = \binom{2m}{m}\binom{m}{m}\frac{m!}{(m+1)(m+2)\ldots\ldots(2m+1)} = \frac{1}{2m+1} \qquad (6.4.9)$$

of $\int_0^1 P_m^2(x)dx = \frac{1}{2m+1}$.

De in deze paragraaf ingevoerde $P_m(x)$ voldoen volgens (6.4.8) en (6.4.9) aan de eisen (6.3.3). Verder is $P_m(0) = 1$. De $P_m(x)$ uit deze paragraaf zijn dus dezelfde als die uit de vorige.

Met de $P_m(x)$ gedefinieerd volgens (6.4.7) vinden wij dus de coëfficiënten b_j volgens (6.3.6).

De polynomen $P_m(x)$ met $0 \leq x \leq 1$ heten polynomen van Legendre. (eigenlijk zijn $P_m(\frac{x+1}{2})$ met $-1 \leq x \leq 1$ als functie van x beschouwd, de polynomen van Legendre).

De eerste vier polynomen zijn:

$$P_0(x) = 1$$
$$P_1(x) = 1 - 2x$$
$$P_2(x) = 1 - 6x + 6x^2$$
$$P_3(x) = 1 - 12x + 30x^2 - 20x^3$$

Eenvoudig is in te zien dat alle nulpunten reeel, enkelvoudig, positief <1 zijn: Neem daartoe de reele nulpunten van oneven orde $x_1 \ldots\ldots x_r$ met $0 < x_i < 1$, waarbij $r \leq m$ is.

Het polynoom $P_m(x)(x - x_1) \ldots\ldots (x - x_r)$ wisselt niet van teken op het interval 0 - 1 en $\int_0^1 P_m(x)(x - x_1) \ldots\ldots (x - x_r)dx$ is dus ongelijk nul. Maar volgens (6.4.1) moet dan $r = m$ zijn; anders was de integraal wel nul. Dus zijn alle nulpunten van $P_m(x)$ reeel, enkelvoudig en positief < 1.

6.5 Kleinste kwadraten approximatie bij discrete equidistante punten.

Zijn de x_i uit § 6.2 equidistant dan zal deze paragraaf een methode geven om het stelsel (6.2.2) op eenvoudige wijze op te lossen. Daartoe wordt gebruik gemaakt van factorpolynomen

$$s^{(n)} = s(s - 1) \ldots\ldots (s - n + 1) \tag{6.5.1}$$

Voor de voorwaartse differentie hiervan vindt men

$$\Delta s^{(n)} = (s + 1)^{(n)} - s^{(n)} = (s + 1)s(s-1)\ldots\ldots(s - n + 2) -$$
$$- s(s - 1)\ldots\ldots (s - n + 1) =$$
$$= s(s - 1) \ldots\ldots (s - n + 2)(s + 1 - s + n - 1) = ns^{(n-1)}$$
$$\therefore \Delta s^{(n)} = ns^{(n - 1)} \tag{6.5.2}$$

Deze formule laat zien dat $s(n)$ zich t.o.v. differenties maken gedraagt als s^n t.o.v. differentieren. De analogie gaat nog verder. Beschouw

$$\sum_{s=M}^{N} \Delta f(s) = f(N + 1) - f(N) + f(N) - f(N - 1) \ldots\ldots - f(M) = f(s)\Big/_{M}^{N+1}.$$

Nemen we nu i.p.v. $f(s)$, $s^{(n+1)}$:

$$\sum_{s=M}^{N} \Delta s^{(n+1)} = (n + 1) \sum_{s=M}^{N} s^{(n)} = s^{(n+1)}\Big/_{M}^{N+1} \quad \text{of}$$

$$\sum_{s=M}^{N} s^{(n)} = \frac{1}{n + 1} s^{(n+1)}\Big/_{M}^{N+1} \tag{6.5.3}$$

Deze formule is het analogon voor $\int_M^N s^n ds = \frac{1}{n + 1} s^{n+1}\Big/_{M}^{N}$

Nu definieren wij polynomen

$$P_{m,n}(x) = 1 + b_1 x + b_2 x^{(s)} + \ldots\ldots + b_m x^{(m)} \tag{6.5.4}$$

met de eigenschap:

$$\sum_{x=0}^{n} P_{m,n}(x)(x+s)^{(s)} = 0 \quad \text{als} \quad s = 0(1)m-1 \qquad (6.5.5)$$

Wij zullen nu de coëfficiënten b_i bepalen. Daartoe substitueren wij (6.5.4) in (6.5.5). Wij vinden dan:

$$\sum_0^n (x+s)^{(s)} + b_1 \sum_0^n (x+s)^{(s+1)} + b_2 \sum_0^n (x+s)^{(s+2)} + \dots$$

$$\dots\ b_m \sum_0^n (x+s)^{(s+m)} = 0 \text{ voor } s = 0(1)m-1 \qquad (6.5.6)$$

Met (6.5.3) vinden wij:

$$\frac{(n+s+1)^{(s+1)}}{s+1} + b_1 \frac{(n+s+1)^{(s+2)}}{s+2} + b_2 \frac{(n+s+1)^{(s+3)}}{s+3} + \dots$$

$$\dots + b_m \frac{(n+s+1)^{(s+m+1)}}{s+m+1} = 0 \text{ voor } s = 0(1)m-1 \ (6.5.7)$$

waarbij opgemerkt dient te worden dat de termen afkomstig van de ondergrens in (6.5.3) nul zijn wegens

$$s^{(n)} = 0 \quad \text{als} \quad n > s$$

hetgeen direct uit (6.5.1) volgt. $(s^{(s)} = s!)$.

Vervolgens delen wij (6.5.7) door $(n+s+1)^{s+1}$ en vinden

$$\frac{1}{s+1} + b_1 \frac{n}{s+2} + b_2 \frac{n^{(2)}}{s+3} + \dots + b_m \frac{n^{(m)}}{s+m+1} = 0 \qquad (6.5.8)$$

Stelt men nu $n^{(k)} b_k = a_k$ dan gaat (6.5.8) over in (6.4.3) waardoor wij direct vinden $b_k = \frac{1}{n^{(k)}} (-)^k \binom{m+k}{k}\binom{m}{k}$ en

$$P_{m,n}(x) = \sum_{k=0}^{m} (-)^k \frac{1}{n^{(k)}} \binom{m+k}{k}\binom{m}{k} x^{(k)} \qquad (6.5.9)$$

Enkele van deze polynomen zijn:

$$P_{0,n}(x) = 1$$
$$P_{1,n}(x) = 1 - 2\frac{x}{n}$$
$$P_{2,n}(x) = 1 - 6\frac{x}{n} + 6\,\frac{x(x-1)}{n(n-1)}$$
$$P_{3,n}(x) = 1 - 12\frac{x}{n} + 30\,\frac{x(x-1)}{n(n-1)} - 20\,\frac{x(x-1)(x-2)}{n(n-1)(n-2)}$$

Om nu met deze $P_{m,n}(x)$ een kleinste kwadraten oplossing te vinden, gaan wij als volgt te werk. Laten de equidistante basis-

punten x_i $(i = 0(1)n)$ vervangen worden door hun rangnummer i.
Wij stellen onze approximatie voor door

$$\sum_{k=0}^{n} c_k P_{k,n}(i) \tag{6.5.10}$$

Voor een kleinste kwadratenoplossing moet dan gelden:

$$\sum_{0}^{n} \left\{ f_i - \sum_{0}^{m} c_k P_{k,n}(i) \right\}^2 = \min$$

of $$\sum_{i=0}^{n} P_{l,n}(i) \sum_{k=0}^{m} c_k P_{k,n}(i) = \sum_{i=0}^{n} P_{l,n}(i) f(i)$$

of $$\sum_{k=0}^{m} c_k \left\{ \sum_{i=0}^{n} P_{l,n}(i) P_{k,n}(i) \right\} = \sum_{i=0}^{n} P_{l,n}(i) f(i) \tag{6.5.11}$$

Nu is natuurlijk $\sum_{i=0}^{n} P_{l,n}(i) P_{k,n}(i) = 0$ als $l \neq k$: (6.5.12)

Laat $k < l$. $P_{k,n}(i)$ is te ontwikkelen in een reeks factorpolynomen:

$$P_{k,n}(i) = \sum_{s=0}^{k} b_s i^{(s)} \tag{6.5.13}$$

Substitueert men (6.5.13) in het linkerlid van (6.5.12) dan vindt men op grond van (6.5.5) de geldigheid van (6.5.12).
In het linkerlid van (6.5.4) blijft dan alleen over
$c_l \sum_{i=0}^{n} P_{l,n}^2(i)$ en wij vinden:

$$c_l = \frac{\sum_{i=0}^{n} P_{l,n}(i) f(i)}{\sum_{i=0}^{n} P_{l,n}^2(i)} . \tag{6.5.14}$$

Op deze wijze vinden wij dan zonder een lineair stelsel op te lossen de c_k. Door daarna de $P_{k,n}(i)$ als polynomen in i te schrijven vinden wij uit (6.5.10) de gevraagde polynoomoplossing.

6.6. Toepassingen

Als toepassing van § 6.5 behandelen wij de opgave van pag. 3:

s	x	y
0	-3	-.71
1	-2	-.01
2	-1	+.51
3	0	+.82
4	1	+.88
5	2	+.81
6	3	+.49

Gevraagd was een 2e graads approximatie te contrueren. Wij voeren in een nieuwe onafhankelijke variabele s = x+3; wij hebben n = 6. De coëfficiënten van de approximatie

$y^x = \sum c_k P_{k,6}(s)$ vinden wij volgens (6.5.14) als

$$c_k = \frac{\sum_0^6 P_{k,6}(s).f(s)}{\sum_0^6 P_{k,6}^2(s)} .$$

Om de 2e graads approximatie te vinden moeten wij dus berekenen

$P_{0,6}(s)$ t/m $P_{2,6}(s)$ voor $s = 0(1)6$.

Nu is $P_{0,6}(s) = 1$; $P_{1,6}(s) = 1 - \frac{1}{3}s$; $P_{2,6}(s) = 1 - s + \frac{1}{5}s(s-1)$.
Dus:

s	P_0	P_1	P_2	P_3
0	1	1	1	1
1	1	2/3	0	-1
2	1	1/3	-3/5	-1
3	1	0	-4/5	0
4	1	-1/3	-3/5	1
5	1	-2/3	0	1
6	1	-1	1	-1

en wij vinden $c_o = .3986$; $c_1 = -.6011$; $c_2 = -.5089$ of

$y^x(s) = .3986 - .6011\ P_1(s) - .5089\ P_2(s) = .806 + .200x - .102x^2$,

hetgeen overeenstemt met het vroeger gevonden antwoord.
Willen wij een 3e graads approximatie maken, dan behoeven wij slechts te bepalen

$$P_{3,6}(s) = 1 - 2s + s(s-1) - \frac{s}{6}(s-1)(s-2)\ .$$

Wij vinden dan $c_3 = -.00167$ en zien dat deze term nagenoeg geen verandering meer oplevert t.o.v. de tweede graads approximatie.

Twee dingen dienen opgemerkt te worden:

1. Bij het gebruik maken van de orthogonale polynomen behoeft men niet zoals bij de in (6.2) geschetste methode al het werk over te doen als men een approximatie van hogere graad wil maken.
2. Een approximatie van hogere graad is nooit slechter dan een van lagere graad: was het wel het geval dan zou men dus een betere approximatie verkrijgen door enkele $c_k = 0$ te stellen. Dit kan niet daar de c_k zo gekozen waren dat ze de afwijkingen minimaliseerden.

6.7. Tabellen

In deze paragraaf geven wij enkele tabellen van $P_{m,n}(x)$.

s	$P_o(s)$	$P_1(s)$	$P_2(s)$	$P_3(s)$	$P_4(s)$	$P_5(s)$
			n = 5			
0	1	5	5	5	1	1
1	1	3	-1	-7	-3	-5
2	1	1	-4	-4	2	10
3	1	-1	-4	4	2	10
4	1	-3	-1	7	-3	-5
5	1	-5	5	-5	1	1
	6	70	84	180	28	252

s	$P_0(s)$	$P_1(s)$	$P_2(s)$	$P_3(s)$	$P_4(s)$	$P_5(s)$
			n = 6			
0	1	3	5	1	3	1
1	1	2	0	-1	-7	-4
2	1	1	-3	-1	1	5
3	1	0	-4	0	6	0
4	1	-1	-3	1	1	-5
5	1	-2	0	1	-7	4
6	1	-3	5	-1	3	-1
	7	28	84	6	154	84
			n = 7			
0	1	7	7	7	7	7
1	1	5	1	-5	-13	-23
2	1	3	-3	-7	- 3	17
3	1	1	-5	-3	9	15
4	1	-1	-5	3	9	-15
5	1	-3	-3	7	- 3	-17
6	1	-5	1	5	-13	23
7	1	-7	7	-7	7	- 7
	8	168	168	264	616	2184
			n = 8			
0	1	4	28	14	14	4
1	1	3	7	- 7	-21	-11
2	1	2	- 8	-13	-11	4
3	1	1	-17	- 9	9	9
4	1	0	-20	0	18	0
5	1	-1	-17	9	9	-9
6	1	-2	- 8	13	-11	-4
7	1	-3	7	7	-21	11
8	1	-4	28	-14	14	-4
	9	60	2772	990	2002	468

Bij het gebruik van deze tabellen dient men te bedenken dat niet is opgegeven $P_{m,n}(s)$ maar een zodanig veelvoud c, dat $P_{m,n}(s)$ voor $s = 0(1)n$ geheel is. Daar $P_{m,n}(0) = 1$ is, kan men de factor c dus aflezen aan de opgegeven waarde van $P_{m,n}(0)$. Verder is onder iedere kolom opgegeven de som der kwadraten van de in die kolom voorkomende getallen: in deze som komt de factor c tweemaal voor. Bij het berekenen van (6.5.14) met behulp van deze tabellen moet men dus met c vermenigvuldigen om de juiste coëfficient c_k te vinden.

6.8. Harmonische Analyse

In de voorgaande paragrafen hebben wij kleinste kwadraten approximaties beschouwd met polynomen. Hebben wij een periodieke functie met periode $2\pi(f(x+k.2\pi) = f(x))$ dan ligt het voor de hand om niet met polynomen maar met trigonometrische sommen te benaderen.
Wij zullen $f(x)$ benaderen op het interval 0-2π door

$$y(x) = a_0 + \sum_1^m a_k \cos kx + \sum_1^m b_k \sin kx \tag{6.8.1}$$

en wel in de zin der kleinste kwadraten, dus zodat

$$F = \int_0^{2\pi} \left\{f(x) - y(x)\right\}^2 dx \text{ minimaal is} \tag{6.8.2}$$

Nul stellen der partiële afgeleiden geeft de relaties:

$$\frac{\partial F}{\partial a_i} = -2\int_0^{2\pi} \left\{f(x) - a_0 - \sum_1^m a_k \cos kx - \sum_1^m b_k \sin kx\right\}. \cos ix \, dx = 0$$

$$\frac{\partial F}{\partial b_i} = -2\int_0^{2\pi} \left\{f(x) - a_0 - \sum_1^m a_k \cos kx - \sum_1^m b_k \sin kx\right\}. \sin ix \, dx = 0$$

$$\frac{\partial F}{\partial a_0} = -2\int_0^{2\pi} \left\{f(x) - a_0 - \sum_1^m a_k \cos kx - \sum_1^m b_k \sin kx\right\} dx = 0$$

(6.8.3)

Voorbeelden kleinste kwadraten approximatie

Gevraagd te approximeren sin $\pi/2$ x $\quad 0 \leq x \leq 1$ in de zin der kleinste kwadraten met een polynoomgraad 2.

Volgens (6.3.6) vindt men voor de approximatie $\sin^* \frac{\pi}{2} x$

$$\sin^* \frac{\pi}{2} x = b_o P_o(x) + b_2 P_2(x)$$

met

$$b_o = \int_o^1 \sin \pi/2 \; x \; dx$$

$$b_1 = 3 \int_o^1 \sin \pi/2 \; x(1-2x)dx$$

$$b_2 = 5 \int_o^1 \sin \pi/2 \; x(1-6x+6x^2)dx$$

Als men deze integralen berekent, vindt men:

$$b_o = 2/\pi \qquad = .6366\ 1977$$

$$b_1 = 6/\pi - 24/\ ^2 \qquad = -.5218\ 4909$$

$$b_2 = 10/\pi + 120/\pi^2 - 480/\pi^3 = -.1390\ 9562$$

Dus $\sin^* \frac{\pi}{2} x = .6366\ 1977\ P_o - .5218\ 4909\ P_1 - .1390\ 9562\ P_2 =$

$= .6366\ 1977.1 - .5218\ 4909(1-2x) - .1390\ 9562(1-6x+6x^2) =$

$= -.02432 + 1.87827\ x - .83457\ x^2$

x	$\sin^* \pi/2$	$\sin \pi/2\ x$	$\sin^* \pi/2\ x - \sin \pi/2$
0	-.02432	0	-.024
.1	.15516	.15643	-.001
.2	.31745	.30902	+.008
.3	.46405	.45399	+.010
.4	.59346	.58779	+.006
.5	.70617	.70711	-.001
.6	.80220	.80902	-.007
.7	.88153	.89101	-.009
.8	.94417	.95106	-.007
.9	.99012	.98769	+.002
1.0	1.01938	1.00000	+.019

Uit (6.8.3) dienen a_k en b_k bepaald te worden. Dit wordt zeer vergemakkelijkt door de orthogonaliteitsrelaties:

$$\left.\begin{aligned}
&\int_0^{2\pi} \cos kx \cos lx \, dx = \begin{cases} 0 & \text{als } k \neq l \\ \pi & \text{als } k = l \\ 2\pi & \text{als } k = l = 0 \end{cases} \\
&\int_0^{2\pi} \sin kx \cos lx \, dx = 0 \\
&\int_0^{2\pi} \sin kx \sin lx \, dx = \begin{cases} 0 & \text{als } k \neq l \\ \pi & \text{als } k = l \end{cases}
\end{aligned}\right\} \qquad (6.8.4)$$

(6.8.4) leidt men als volgt af:

$$\int_0^{2\pi} \cos kx \cos lx \, dx = \tfrac{1}{2} \int_0^{2\pi} \left\{ \cos(k+l)x + \cos(k-l)x \right\} dx = 0 \text{ als } k \neq l.$$

Als $k = l$ dan:

$$\tfrac{1}{2} \int_0^{2\pi} \left\{ \cos 2kx + 1 \right\} dx = \pi$$

$$\int_0^{2\pi} \sin kx \cos lx \, dx = \tfrac{1}{2} \int_0^{2\pi} \left\{ \sin(k+l)x + \sin(k-l)x \right\} dx = 0$$

en tenslotte $\displaystyle\int_0^{2\pi} \sin kx \sin lx \, dx = \tfrac{1}{2} \int_0^{2\pi} \left\{ \cos(k-l)x - \cos(k+l)x \right\} dx = 0$ als $k \neq l$

en als $k = l \neq 0$ dan: $\displaystyle\tfrac{1}{2} \int_0^{2\pi} \left\{ 1 - \cos 2kx \right\} dx = \pi$

Met behulp van (6.8.4) vindt men de oplossing van (6.8.3):

$$\left.\begin{aligned}
a_0 &= \frac{1}{2\pi} \int_0^{2\pi} f(x) \, dx \\
a_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx \\
b_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx
\end{aligned}\right\} \qquad (6.8.5)$$

In deze uitdrukkingen komt m niet voor: a_k en b_k hangen blijkbaar alleen af van $f(x)$ en van k: men kan ze dus onafhankelijk van elkaar berekenen.

6.9. Harmonische Analyse (equidistante discrete punten)

Is de periodieke functie $f(x)$ alleen gegeven op n discrete basispunten $x_l = \frac{2\pi l}{n}$, $l = 1(1)n$ dan eisen wij in plaats van (6.8.2) dat

$$F = \sum_{l=1}^{n} \{f(x) - y(x_l)\}^2 \text{ minimaal is .} \qquad (6.9.1)$$

In plaats van (6.8.3) komt nu:

$$\left.\begin{aligned}
&\sum_{l=1}^{n} \{f(x_l) - a_o - \sum_{1}^{m} a_k \cos kx_l - \sum_{1}^{m} b_k \sin kx_l\} \cos ix_l = 0 \\
&\sum_{l=1}^{n} \{f(x_l) - a_o - \sum_{1}^{m} a_k \cos kx_l - \sum_{1}^{m} b_k \sin kx_l\} \sin ix_l = 0 \\
&\sum_{l=1}^{n} \{f(x_l) - a_o - \sum_{1}^{m} a_k \cos kx_l - \sum_{1}^{m} b_k \sin kx_l\} = 0
\end{aligned}\right\} (6.9.2)$$

In (6.9.2) komen voor (als men de sommaties verwisselt) sommen van het type:

$$\left.\begin{aligned}
&\sum_{l=1}^{n} \cos\frac{2\pi l}{n}k \cos\frac{2\pi l}{n}i \\
&\sum_{l=1}^{n} \cos\frac{2\pi l}{n}k \sin\frac{2\pi l}{n}i \\
&\sum_{l=1}^{n} \sin\frac{2\pi l}{n}k \sin\frac{2\pi l}{n}i \\
&\sum_{l=1}^{n} \cos\frac{2\pi l}{n}k \text{ en } \sum_{l=1}^{n} \sin\frac{2\pi l}{n}k \text{ .}
\end{aligned}\right\} \qquad (6.9.3)$$

Deze uitdrukkingen zijn gemakkelijk te sommeren. Wij beginnen met de laatste twee.

$$\cos\frac{2\pi l}{n}k = \mathrm{Re}\; e^{\frac{2\pi k}{n}li} \quad \text{en} \quad \sin\frac{2\pi l}{n}k = \mathrm{Im}\; e^{\frac{2\pi k}{n}li} \text{ .}$$

Wij vinden dan

$$\sum_{l=1}^{n} (\cos\frac{2\pi l}{n}k + i\sin\frac{2\pi l}{n}k) = \sum_{l=1}^{n} e^{\frac{2\pi k l}{n}i} .$$

De reeks rechts is een meetkundige en wel met reden $e^{\frac{2\pi k}{n}i}$.
De som is dus:

$$e^{\frac{2\pi k}{n}i} \frac{1 - e^{\frac{2\pi k n}{n}i}}{1 - e^{\frac{2\pi k}{n}i}} = 0 \quad \text{tenzij } k = 0 .$$

Dus geldt $\sum_{l=1}^{n} \cos\frac{2\pi k}{n}l = 0$ en $\sum_{l=1}^{n} \sin\frac{2\pi k}{n}l = 0$ als $k \neq 0$ (6.9.4)

Als $k = 0$ dan geldt $\sum_{l=1}^{n} \cos\frac{2\pi k}{n}l = n$ en $\sum_{l=1}^{n} \sin\frac{2\pi k}{n}l = 0$.

Verder geldt:

$$\sum_{l=1}^{n} \cos\frac{2\pi k}{n}l \cos\frac{2\pi i}{n}l = \tfrac{1}{2} \sum_{l=1}^{n} \left\{\cos\frac{2\pi l}{n}(k+i) + \cos\frac{2\pi l}{n}(k-i)\right\}$$

en dit is nul volgens (6.9.4) tenzij $k=i$; in dat geval is de som $\frac{1}{2}n$. Geldt $k=i=0$ dan is de som n.

Op gelijke wijze vindt men $\sum_{1}^{n} \cos\frac{2\pi l}{n}k \sin\frac{2\pi l}{n}i = 0$

en
$$\sum_{l=1}^{n} \sin\frac{2\pi l}{n}k \sin\frac{2\pi l}{n}i \begin{cases} = 0 & \text{als } k \neq i \\ = \frac{n}{2} & \text{als } k = i . \end{cases}$$

Met deze resultaten gaat (6.9.2) over in:

$$\left.\begin{aligned} a_i &= \frac{2}{n} \sum_{l=1}^{n} f(\frac{2\pi l}{n}) \cos\frac{2\pi l}{n}i \\ b_i &= \frac{2}{n} \sum_{l=1}^{n} f(\frac{2\pi l}{n}) \sin\frac{2\pi l}{n}i \\ a_o &= \frac{1}{n} \sum_{l=1}^{n} f(\frac{2\pi l}{n}) . \end{aligned}\right\} \qquad (6.9.5)$$

Het is duidelijk dat het aantal coëfficiënten a_k en b_k dat bepaald kan worden maximaal gelijk is aan het aantal basispunten n.

Is het aantal basispunten oneven, $n = 2r + 1$, dan kan men bepalen a_o t/m a_r en b_1 t/m b_r volgens (6.9.5). Is n even, $n = 2r$, dan kan men bepalen a_o t/m a_r, b_1 t/m b_{r-1}, waarbij men voor a_r de helft van de waarde verkregen volgens (6.9.5) moet nemen:

Met $n = 2r$ en $i = n$ geeft (6.9.2):

$$\sum_{l=1}^{2r} \left\{ f(x_l) - a_o - \sum_1^r a_k \cos\frac{2\pi kl}{2r} - \sum_1^r b_k \sin\frac{2\pi kl}{2r} \right\} \cos\frac{2\pi rl}{2r} = 0$$

De coëfficiënt van a_r wordt $\sum\limits_{l=1}^{2r} \cos\frac{2\pi rl}{2r} \cos\frac{2\pi rl}{2r} = 2r$.

Dus $a_r = \frac{1}{2r} \sum\limits_{l=1}^{2r} f(x_l) \cos\frac{2\pi rl}{2r} = \frac{1}{2r} \sum\limits_{l=1}^{2r} (-)^l f(x_l)$.

6.10 Numerieke bepaling van Fourier-coëfficiënten

Een veel voorkomende opgave is om voor een gegeven periodieke functie met periode 2π

$$f(x) = a_o + \sum_1^\infty a_k \cos kx + \sum_1^\infty b_k \sin kx \tag{6.10.1}$$

te bepalen $a_o, a_1, a_2, \ldots$ en $b_1, b_2, \ldots$.

Men vindt:
$$\begin{aligned} a_o &= \frac{1}{2\pi}\int_0^{2\pi} f(x)dx \\ a_k &= \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\, dx \\ b_k &= \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\, dx \end{aligned} \tag{6.10.2}$$

Immers vermenigvuldig (6.10.1) met $\genfrac{}{}{0pt}{}{\cos}{\sin}kx$ en integreer. (6.10.2) blijkt identiek met (6.8.5).

Is f(x) alleen gegeven in $x_l = \frac{2\pi l}{n}$, $l = 1(1)n$, dan zal men de (6.10.2) numeriek moeten bepalen. Toepassing van de trapeziumregel geeft, rekening houdend met $f(0) = f(2\pi)$:

$$a_o^* = \frac{1}{n} \sum_1^n f\left(\frac{2\pi l}{n}\right)$$

$$a_k^* = \frac{2}{n} \sum_1^n f\left(\frac{2\pi l}{n}\right) \cos\frac{2\pi kl}{n} \qquad (6.10.3)$$

$$b_k^* = \frac{2}{n} \sum_1^n f\left(\frac{2\pi l}{n}\right) \sin\frac{2\pi kl}{n}$$

(6.10.3) blijkt identiek te zijn met (6.9.5).
Men kan nu in de verleiding komen de integralen in (6.10.2) met een betere integratieformule te berekenen, b.v. Simpson; dit heeft echter geen zin zoals o.a. uit (4.4.2) volgt: de differentiecorrecties vallen weg. Dit wil niet zeggen dat het antwoord exact is; verandering van het aantal basispunten geeft een ander antwoord. Door in rekening brengen van differentiecorrecties kan de fout echter niet beïnvloed worden.

Wij zullen nu laten zien welke fout men maakt door (6.10.3) als benadering van (6.10.2) te gebruiken. Daartoe substitueren wij (6.10.1) in (6.10.3):

$$a_o^* = \frac{1}{n} \sum_1^n f\left(\frac{2\pi l}{n}\right) = \frac{1}{n} \sum_1^n \left\{a_o + \sum_1^\infty a_k \cos\frac{2\pi kl}{n} + \sum_1^\infty b_k \sin\frac{2\pi kl}{n}\right\} =$$

$$= a_o + \frac{1}{n} \sum_1^\infty a_k \left(\sum_1^n \cos\frac{2\pi kl}{n}\right) + \frac{1}{n} \sum_1^\infty b_k \left(\sum_1^n \sin\frac{2\pi kl}{n}\right) =$$

$$= a_o + a_n + a_{2n} + \ldots$$

$$a_k^* = \frac{2}{n} \sum_1^n f\left(\frac{2\pi l}{n}\right) \cos\frac{2\pi kl}{n} = \frac{2}{n} \sum_1^n \left\{a_o + \sum_1^\infty a_m \cos\frac{2\pi ml}{n} + \sum_1^\infty b_m \sin\frac{2\pi ml}{n}\right\} . \cos\frac{2\pi kl}{n} =$$

$$= \frac{2}{n} \sum_1^\infty a_m \sum_1^n \cos\frac{2\pi ml}{n} \cos\frac{2\pi kl}{n} = \frac{1}{n} \sum_1^\infty a_m \sum_1^n \left\{\cos\frac{2\pi l}{n}(m-k) + \cos\frac{2\pi l}{n}(m+k)\right\} =$$

$$= a_k + a_{n+k} + a_{2n+k} + a_{3n+k} + \ldots$$

$$+ a_{n-k} + a_{2n-k} + a_{3n-k} + \ldots$$

$$b_k^* = \frac{2}{n}\sum_1^n f\left(\frac{2\pi l}{n}\right) \sin\frac{2\pi kl}{n} = \frac{2}{n}\sum_1^n \left\{a_o + \sum_1^\infty a_m \cos\frac{2\pi ml}{n} + \sum_1^\infty b_m \sin\frac{2\pi ml}{n}\right\} . \sin\frac{2\pi kl}{n} =$$

$$= \frac{2}{n}\sum_1^\infty b_m \sum_1^n \sin\frac{2\pi ml}{n} \sin\frac{2\pi kl}{n} = \frac{1}{n}\sum_1^\infty b_m \sum_1^n \left\{\cos\frac{2\pi l}{n}(m-k) \quad \cos\frac{2\pi l}{n}(m+k)\right\} =$$

$$= b_k + b_{n+k} + b_{2n+k} + b_{3n+k} + \dots$$

$$- b_{n-k} - b_{2n-k} - b_{3n+k} - \dots .$$

Wij vinden dus voor de numerieke Fourier-coëfficiënten:

$$\left.\begin{aligned} a_o^* &= a_o + a_n + a_{2n} + a_{3n} + \dots \\ a_k^* &= a_k + a_{n+k} + a_{2n+k} + a_{3n+k} + \dots \\ &\quad + a_{n-k} + a_{2n-k} + a_{3n-k} + \dots \\ b_k^* &= b_k + b_{n+k} + b_{2n+k} + b_{3n+k} + \dots \\ &\quad - b_{n-k} - b_{2n-k} - b_{3n-k} - \dots \end{aligned}\right\} \qquad (6.10.4)$$

Hieruit volgt dus de fout die wij maken bij de numerieke integratie. De fout hangt dus sterk af van de convergentie van de Fourierreeks en groeit met toenemende k (als n vast is). Voor $k \approx \frac{1}{2}n$ is de fout ongeveer 100% nl. $a^*_{n/2} \approx 2a_{n/2}$.

Uit (6.10.4) volgt dat het vrijwel niet mogelijk is zonder kennis van de convergentie van de echte Fourier-coëfficiënten een k aan te geven waarvoor men bij gegeven n de a_k en b_k met enige precisie kan benaderen. Omtrent de fout kan men natuurlijk iets te weten komen door ook met een ander aantal punten te rekenen. Daarbij moet men wel oppassen: verdubbelt men het aantal n en had men te maken met een functie waarvoor toevallig $a_n = 0$ dan was dus in het geval van n punten de fout in a_o ongeveer a_{2n}; met 2n punten vindt men ook een fout a_{2n}: men vindt dus geen verschil tussen beide benaderingen van a_o en zou tot hoge precisie concluderen; dit hoeft geenszins het geval te zijn. In het algemeen is het beter de berekening te doen voor twee waarden van n die relatief priem zijn.

Voorbeeld

De Fourier-ontwikkeling van cos(cosx) heeft, daar de functie even is alleen cos-termen:

$$\cos(\cos x) = a_o + \sum_1^{\infty} a_k \cos kx$$

met

$$a_k = \frac{1}{\pi}\int_0^{2\pi} \cos(\cos x)\cos kx\, dx = \frac{2}{\pi}\int_0^{\pi} \cos(\cos x)\cos kx\, dx =$$

$$= \frac{2}{\pi}\left[\int_0^{\frac{\pi}{2}} \cos(\cos x)\cos kx\, dx + \int_{\frac{\pi}{2}}^{\pi} \cos(\cos x)\cos kx\, dx\right].$$

Substitueer in de tweede integraal $x = \pi - v$:

$$= \frac{2}{\pi}\left[\int_0^{\frac{\pi}{2}} \cos(\cos x)\cos kx\, dx + (-)^k\int_0^{\frac{\pi}{2}} \cos(\cos v)\cos kv\, dv\right] = 0$$

als k oneven.

$\therefore\ a_{2k+1} = 0.$

Neemt men $n = 2r + 1$, dan is de fout in a_o ongeveer $a_{2(2r+1)}$; voor $n = 4r + 2$ vindt men als fout eveneens $a_{2(2r+1)}$.

6.11. Polynomen van Chebyshev

De stelling van Chebyshev luidt:

Onder de polynomen $p_r(x)$ van de graad n met coëfficiënt van x^n gelijk aan 1 is er een en slechts een waarvoor op het interval $-1 \leq x \leq 1$ het maximum van $|p_n(x)|$ minimaal is.
Dit polynoom is

$$2^{-(n-1)} T_n(x)\,. \tag{6.11.1}$$

$$\text{waarbij } T_n(x) = \cos(n\ \mathrm{bgcos}x)\,. \tag{6.11.2}$$

Om in te zien dat (6.11.2) een polynoom in x is stellen wij

$$\begin{aligned} &\mathrm{bg}\cos x = \Theta \\ &\text{of} \\ &x = \cos\Theta \end{aligned} \tag{6.11.3}$$

(6.11.2) gaat dan over in $T_n(\cos\theta) = \cos n\theta$. (6.11.4)

Maar $\cos n\theta$ is uit te drukken als polynoom in $\cos\theta$:

$$\cos n\theta = \mathrm{Re}(\cos\theta + i\sin\theta)^n = \mathrm{Re}\sum_0^n i^k\binom{n}{k}\cos^{n-k}\theta\,\sin^k\theta =$$

$$= \binom{n}{0}\cos^n\theta - \binom{n}{2}\cos^{n-2}\theta\,\sin^2\theta + \binom{n}{4}\cos^{n-4}\theta\,\sin^4\theta - \ldots$$

Tenslotte geeft substitutie van $\sin^2\theta = 1 - \cos^2\theta$ het gewenste polynoom in $\cos\theta$. De coëfficiënt van $\cos^n\theta$ is gelijk aan:

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots = 2^{n-1} \qquad (6.11.5)$$

Wij hebben dus aangetoond, dat $T_n(x)$ een polynoom is in x van de graad n met 2^{n-1} als coëfficiënt van x^n; $2^{-(n-1)}T_n(x)$ heeft dus 1 als coëfficiënt van x^n. De maximale absolute waarde is $2^{-(n-1)}$ en deze wordt aangenomen in de punten:

$$x_k = \cos k\frac{\pi}{n} \qquad k = 0(1)n \qquad (6.11.6)$$

Men kan nu bewijzen dat er geen polynoom van de gewenste vorm is met kleinere maximale waarde: $2^{-(n-1)}T_n(x)$ is het gezochte minimaal-polynoom.

Wij weten: $\cos(n+1)\theta + \cos(n-1)\theta = 2\cos n\theta\cos\theta$.

Dit, vertaald in $T_n(\cos\theta)$, geeft:

$$T_{n+1}(\cos\theta) + T_{n-1}(\cos\theta) = 2T_n(\cos\theta)\cos\theta$$

of met de variabele x:

$$T_{n+1}(x) - 2xT_n(x) + T_{n-1}(x) = 0 \qquad (6.11.7)$$

Dit is de recursie-betrekking voor $T_n(x)$. De startwaarden $T_0(x)$ en $T_1(x)$ vinden wij uit (6.11.4):

$$T_0(x) = 1$$

$$T_1(x) = x$$

Met (6.11.7) kunnen wij dus opvolgende $T_n(x)$ bepalen.

$$
\begin{aligned}
T(x) &= 1 \\
T(x) &= x \\
T(x) &= 2x^2 - 1 \\
T(x) &= 4x^3 - 3x \\
T(x) &= 8x^4 - 8x^2 + 1 \\
T(x) &= 16x^5 - 20x^3 + 5x \\
T(x) &= 32x^6 - 48x^4 + 18x^2 - 1 \\
T(x) &= 64x^7 - 112x^5 + 56x^3 - 7x \\
T(x) &= 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1 \\
T(x) &= 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x \\
T(x) &= 512x^{10} - 1280x^8 + 1120x^6 - 400x^4 + 50x^2 - 1
\end{aligned}
\qquad (6.11.8)
$$

Omgekeerd kunnen wij x^n uitdrukken in een lineaire combinatie van $T_o(x)$ t/m $T_n(x)$:

$$
\begin{aligned}
1 &= T_o \\
x &= T_1 \\
x^2 &= \tfrac{1}{2}(T_o + T_2) \\
x^3 &= \tfrac{1}{4}(3T_1 + T_3) \\
x^4 &= \tfrac{1}{8}(3T_o + 4T_2 + T_4) \\
x^5 &= \tfrac{1}{16}(10T_1 + 5T_3 + T_5) \\
x^6 &= \tfrac{1}{32}(10T_o + 15T_2 + 6T_4 + T_6) \\
x^7 &= \tfrac{1}{64}(35T_1 + 21T_3 + 7T_5 + T_7) \\
x^8 &= \tfrac{1}{128}(35T_o + 56T_2 + 28T_4 + 8T_6 + T_8) \\
x^9 &= \tfrac{1}{256}(126T_1 + 84T_3 + 36T_5 + 9T_7 + T_9) \\
x^{10} &= \tfrac{1}{512}(126T_o + 210T_2 + 120T_4 + 45T_6 + 10T_8 + T_{10})
\end{aligned}
\qquad (6.11.9)
$$

Behalve de recursiebetrekking (6.11.7) geven wij nog enkele relaties:

$$\cos m\theta \cos n\theta = \tfrac{1}{2} \cos(m+n)\theta + \cos(m-n)\theta$$

of

$$T_m(x)T_n(x) = \tfrac{1}{2}\{T_{m+n}(x) + T_{m-n}(x)\} = \tfrac{1}{2}\{T_{m+n}(x) + T_{n-m}(x)\} \;. \qquad (6.11.10)$$

Verder geldt:

$$
\begin{aligned}
\int_0^{\pi} \cos j\theta \cos k\theta \, d\theta &= 0 && j \neq k \\
&= \tfrac{\pi}{2} && j = k \neq 0 \\
&= \pi && j = k = 0
\end{aligned}
$$

waaruit volgt door substitutie van $x = \cos\theta$:

$$\int_{-1}^{1} \frac{T_j(x)T_k(x)}{\sqrt{1-x^2}}\,dx = 0 \qquad j \neq k$$
$$= \frac{\pi}{2} \qquad j = k \neq 0 \qquad (6.11.11)$$
$$= \pi \qquad j = k = 0$$

(6.11.11) geeft de orthogonaliteitsrelatie van $T_n(x)$.

Tenslotte hebben wij de relaties:

$$\sum_0^{n-1} \cos j\theta_\alpha \;\cos k\theta_\alpha = 0 \qquad j,k < n, \; j \neq k$$
$$\frac{n}{2} \qquad j = k \neq 0$$

waarbij $\theta_\alpha = \frac{\pi}{2n}(2\alpha + 1)$,

hetgeen oplevert:

$$\sum_0^{n-1} T_j(x_\alpha)T_k(x_\alpha) = 0 \qquad j,k < n, \; j \neq k$$
$$= \frac{n}{2} \qquad j = k \neq 0 \;. \qquad (6.11.12)$$

6.12. Toepassing van Chebyshev polynomen

Hebben wij een functie $f(x)$ $(-1 \leq x \leq 1)$ als machtreeks gegeven, dan kan men met (6.11.9) deze omschrijven in een reeks:

$$f(x) = \sum a_k x^k = \sum b_k T_k(x) \tag{6.12.1}$$

De tweede reeks heeft de eigenschap sneller te convergeren dan de machtreeks. Wil men nu $f(x)$ approximeren met een polynoom met een zekere gegeven precisie, dan heeft men van de machtreeks meer termen nodig dan van de Chebyshev-reeks. Stel men heeft nodig van de machtreeks n, en van de Chebyshev-reeks m termen, dus de approximaties

$$f^{(1)}(x) = \sum_0^n a_k x^k, \quad f^{(2)}(x) = \sum_0^m b_k T_k(x) \ , \tag{6.12.2}$$

dan gaat het tweede polynoom met (6.11.8) over in

$$f^{(2)}(x) = \sum_0^m c_k x^k \ . \tag{6.12.3}$$

De fout is

$$f(x) - f^{(2)}(x) = \sum_{m+1} b_k T_k(x). \tag{6.12.4}$$

Als de reeks $b_k T_k(x)$ snel convergeert is de fout dus ongeveer $b_{m+1} T_{m+1}(x)$,

d.w.z. de fout slingert met maximale absolute waarde b_{m+1}.

Bij deze afgebroken Chebyshev-reeksen krijgt men dus een approximatie, die de eigenschap heeft, dat de fout ongeveer gelijke extremen heeft. Wij zouden liever een approximatie hebben die het maximale aantal extremen heeft, die alle alternerend in teken en exact gelijke absolute waarde hebben. Een dergelijke approximatie bestaat als gevolg van de theorie van Chebyshev, maar numeriek is het niet eenvoudig hem te vinden. De afgebroken Chebyshev-reeks voldoet in de practijk bijna altijd.

Niet altijd is het mogelijk om met (6.11.9) de coëfficiënten b_k te vinden, b.v. als de reeks a_k te slecht convergeert: men zou dan zeer veel termen $a_k x^k$ moeten omschrijven. Dan kan men als volgt te werk gaan:

$$f(x) = \sum b_k T_k(x)$$

Stel $x = \cos\theta$, dan hebben wij:

$$f(\cos\theta) = \sum b_k \cos k\theta, \qquad (6.12.5)$$

of na vermenigvuldiging met $\cos n\theta$ en integratie van θ - π:

$$b_k = \frac{2}{\pi}\int_0^{\pi} f(\cos\theta)\cos k\theta \, d\theta \qquad k \neq 0$$
$$b_0 = \frac{1}{\pi}\int_0^{\pi} f(\cos\theta)d\theta \; . \qquad (6.12.6)$$

In het algemeen krijgt men vrij ingewikkelde integralen uit te rekenen en worden de b_k functies die een klasse moeilijker zijn dan $f(x)$: Is $f(x)$ een sin of e-macht, dan worden b_k Besselfuncties.

Heeft men de gegeven functie in numerieke vorm gegeven voor $-1 \leqslant x \leqslant 1$, dan kan men (door interpolatie b.v.) bepalen:

$$f(x_\alpha) = f(\cos\theta_\alpha) = f\left(\cos\frac{\pi}{2n}(2\alpha+1)\right) \; , \quad \alpha = 0(1)n-1 \quad (6.12.7)$$

waarbij n het tevoren gegeven aantal coëfficiënten is dat men wil bepalen. De fout is dan

$$f(x) - f^{(1)}(x) = \sum_{k=n} b_k T_k(x) \; .$$

Als de reeks snel genoeg convergeert, is dit ongeveer $b_n T_n(x)$. Nu zijn x_α juist de nulpunten van $T_n(x)$. Wij bepalen nu de b_k $k = 0(1)n-1$ zo, dat de approximatie $f^{(1)}(x_\alpha) = f(x_\alpha)$ $\alpha = 0(1)n-1$.

$$\therefore \; f(x_\alpha) = \sum_{k=0}^{n-1} b_k T_k(x_\alpha) \qquad \alpha = 0(1)n-1 \qquad (6.12.8)$$

Om hieruit b_k te bepalen vermenigvuldigen wij met $T_j(x_\alpha)$ en sommeren over α van 0 - (n+1)

$$\sum_{\alpha=0}^{n-1} f(x_\alpha)T_j(x_\alpha) = \sum_{k=0}^{n-1} b_k \sum_{\alpha=0}^{n-1} T_k(x_\alpha)T_j(x_\alpha) \quad . \tag{6.12.9}$$

Volgens (6.11.12) vinden wij

$$b_k = \frac{2}{n}\sum_{\alpha=0}^{n-1} f(x_\alpha)T_k(x_\alpha) \ , \tag{6.12.10}$$

of ook

$$b_k = \frac{2}{n}\sum_{\alpha=0}^{n-1} f(x_\alpha) \cos\frac{\pi k}{2n}(2\alpha+1)$$

$$b_0 = \frac{1}{n}\sum_{\alpha=0}^{n-1} f(x_\alpha) \tag{6.12.11}$$

met

$$x_\alpha = \cos\frac{\pi}{2n}(2\alpha+1) \ .$$

6.13. Voorbeeld

$$f(x) = 1 + \frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{8}x^3 + \frac{1}{16}x^4 \ .$$

Gevraagd 3^e graads approximatie. Wij vinden:

$$f(x) = \frac{147}{128}T_0 + \frac{19}{32}T_1 + \frac{5}{32}T_2 + \frac{1}{32}T_3 + \frac{1}{128}T_4 \ .$$

De approximatie $p_3(x)$ wordt dus

$$p_3(x) = \frac{147}{128}T_0 + \frac{19}{32}T_1 + \frac{5}{32}T_2 + \frac{1}{32}T_3 = \frac{127}{128} + \frac{1}{2}x + \frac{5}{16}x^2 + \frac{1}{8}x^3 \ .$$

Nu met behulp van (6.12.11):

$n = 4$

α	$x_\alpha = \cos\frac{\pi}{8}(2\alpha+1)$ x_α	$\cos\frac{2\pi}{8}(2\alpha+1)$	$\cos\frac{3\pi}{8}(2\alpha+1)$	$f(x_\alpha)$
0	.923880	.707107	.382683	1.819436
1	.382683	- .707107	- .923880	1.236299
2	- .382683	- .707107	.923880	.839606
3	- .923880	.707107	- .382683	.698411

en $b_0 = 1.148438$
$b_1 = .593750$
$b_2 = .156250$
$b_3 = .062500$

hetgeen in overeenstemming is met de eerste antwoorden.

7. Lineaire stelsels

7.1. Inleiding

In dit hoofdstuk zullen enkele directe en indirecte methoden tot het oplossen van lineaire stelsels vergelijkingen behandeld worden. Onder een directe methode verstaan wij een methode, die in een eindig (van tevoren te bepalen) aantal bewerkingen het juiste antwoord levert, mits van afrondingsfouten geen sprake is. Indirecte methoden zijn b.v. iteratieve methoden.

7.2. Eliminatie-methode van Gauss

Laat het lineaire stelsel gegeven zijn als volgt:

$$
\begin{array}{ll}
(1): & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\
(2): & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\
\cdot & \qquad \dots \\
\cdot & \qquad \dots \\
(n): & a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \quad ,
\end{array}
\qquad (7.2.1)
$$

of in matrixvorm $Ax = b$. Wij veronderstellen $\det.A \neq 0$.
Men bepaalt $(2)' = (2) - \frac{a_{21}}{a_{11}}(1)$; $(2)'$ bevat x_1 niet meer.
Daarna $(3)' = (3) - \frac{a_{31}}{a_{11}}(1)$ enz.
Men heeft dan het stelsel:

$$
\begin{array}{l}
a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\
\qquad a^{(1)}_{22}x_2 + \dots + a^{(1)}_{2n}x_n = b^{(1)}_2 \\
\qquad a^{(1)}_{33}x_2 + \dots + a^{(1)}_{3n}x_n = b^{(1)}_3 \\
\qquad\qquad \dots \\
\qquad\qquad \dots \\
\qquad a^{(1)}_{n2}x_2 + \dots + a^{(1)}_{nn}x_n = b^{(1)}_n \quad .
\end{array}
\qquad (7.2.2)
$$

Uit (7.2.2) elimineert men x_2 uit de 3^e t/m n^e vergelijking. Zo voortgaande krijgt men na n-1 eliminaties een stelsel:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \quad \ldots \quad + a_{1n}x_n &= b_1 \\
a_{22}^{(1)}x_2 + \quad \ldots \quad + a_{2n}^{(1)}x_n &= b_2^{(1)} \\
+ a_{33}^{(2)}x_3 + a_{3n}^{(2)}x_n &= b_3^{(2)} \qquad (7.2.3) \\
\vdots \qquad\qquad &\quad \vdots \\
a_{nn}^{(n-1)}x_n &= b_n^{(n-1)}
\end{aligned}
$$

Hieruit bepaalt men achtereenvolgens x_n, $x_{n-1}, \ldots, x_1$.

Men kan ook x_2 elimineren behalve uit de 3^e t/m n^e vergelijking ook uit de 1^e. Doet men dit steeds, dan krijgt men een diagonaalstelsel:

$$
\begin{aligned}
a_{11}^{(n-1)}x_1 \qquad \ldots \qquad &= b_1^{(n-1)} \\
a_{22}^{(n-1)}x_2 \qquad\qquad &= b_2^{(n-1)} \\
\ddots \qquad\qquad &\quad \vdots \qquad (7.2.4) \\
a_{nn}^{(n-1)}x_n &= b_n^{(n-1)}
\end{aligned}
$$

Dit proces heet eliminatie volgens Gauss-Jordan.

Ook voor het inverteren van een matrix A kan men de eleminatiemethode gebruiken. Het gaat dan om de bepaling van B, zodat

AB = I, waarbij I de eenheidsmatrix voorstelt.

De eerste kolom van AB is:

$$
\begin{aligned}
&a_{11}b_{11} + a_{12}b_{21} + \ldots + a_{1n}b_{n1} \\
&a_{21}b_{11} + a_{22}b_{21} + \ldots + a_{2n}b_{n1} \\
&\vdots \\
&a_{n1}b_{11} + a_{n2}b_{21} + \ldots + a_{nn}b_{n1} \,.
\end{aligned}
$$

Opdat dit gelijk is aan de 1^e kolom van I moet gelden:

$$
\begin{aligned}
a_{11}b_{11} + a_{12}b_{21} + \ldots + a_{1n}b_{n1} &= 1 \\
a_{21}b_{11} + a_{22}b_{21} + \ldots + a_{2n}n_{n1} &= 0 \\
\vdots \qquad\qquad\qquad &\vdots \\
a_{n1}b_{11} + a_{n2}b_{21} + \ldots + a_{nn}b_{n1} &= 0
\end{aligned}
\qquad (7.2.5)
$$

De eerste kolom van de inverse A^{-1} vindt men dus door (7.2.5) op te lossen. Door als rechterlid te nemen nullen met op de i^e plaats een 1 vindt men de i^e kolom van A^{-1}.
Inversie komt dus neer op het simultaan oplossen van n stelsels vergelijkingen die een gemeenschappelijke coëfficiëntenmatrix hebben. Dit kan heel geschikt met de eliminatie-methode gedaan worden.

Een andere methode voor het oplossen van vergelijkingen is de regel van Cramer; deze geeft de oplossing van (7.2.1) in determinantvorm als:

$$x_i = \frac{D_i}{D} \qquad (7.2.6)$$

waarbij D = det A en D_i de determinant van de matrix die men krijgt door in A de i^e kolom door de rechterleden b_k te vervangen. Het berekenen van de determinanten vereist zeer veel werk, tenzij men deze weer met de eliminatie-methode berekent. Tenzij het aantal vergelijkingen 2 of 3 is, is deze methode niet aanbevelenswaardig.

Evenmin zal men de inverse van een matrix bepalen door alle minoren uit te rekenen.

7.3. Methode van Crout

De methode van Crout is ook een eliminatie-methode die vooral daarin van die van Gauss verschilt dat het opschrijven van tussenresultaten zeer wordt beperkt. Men hoeft slechts één n x n+1 matrix als hulpgrootheid op te schrijven. Wij gaan weer uit van het stelsel (7.2.1):

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\
\vdots \qquad\qquad\qquad & \quad \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n
\end{aligned} \tag{7.3.1}$$

Wij beschouwen nu de n × n+1 matrix:

$$\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1n} & b_1 \\
a_{21} & a_{22} & & a_{2n} & b_2 \\
\vdots & & & \vdots & \\
a_{n1} & a_{n2} & \dots & a_{nn} & b_n
\end{pmatrix} \tag{7.3.2}$$

en bepalen hiermede een tweede matrix:

$$\begin{pmatrix}
a'_{11} & a'_{12} & \dots & a'_{1n} & b'_1 \\
a'_{21} & a'_{22} & \dots & a'_{2n} & b'_2 \\
\vdots & & & \vdots & \\
a'_{n1} & a'_{n2} & \dots & a'_{nn} & b'_n
\end{pmatrix} \tag{7.3.3}$$

volgens de formules:

$$\begin{aligned}
a'_{ij} &= a_{ij} - \sum_{k=1}^{j-1} a'_{ik}\, a'_{kj} \qquad i \geqslant j \qquad &\text{a)} \\
a'_{ij} &= \frac{1}{a'_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} a'_{ik}\, a'_{kj}\right) \quad i < j \qquad &\text{b)} \\
b'_i &= \frac{1}{a'_{ii}} \left(b_i - \sum_{k=1}^{i-1} a'_{ik}\, b'_k\right) \qquad &\text{c)}
\end{aligned} \tag{7.3.4}$$

De bewering is dat de x_i volgen uit:

$$x_i = b'_i - \sum_{k=i+1}^{n} a'_{ik}\, x_k \tag{7.3.5}$$

Alvorens te bewijzen, dat (7.3.5) de oplossing van (7.3.1) geeft, zullen wij eerst de gang van de berekening schetsen. Men begint volgens (7.3.4a) met het bepalen van a'_{i1}. Wegens $a'_{i1} = a_{i1}$ krijgt men de eerste kolom van (7.3.3) door die van (7.3.2) te copiëren. Vervolgens bepaalt men de rest van de eerste rij:

$$a'_{1j} = \frac{1}{a'_{11}} a_{1j} , \qquad b'_1 = \frac{1}{a'_{11}} b_1$$

Nu kan men de tweede kolom completeren:

$$a'_{i2} = a_{i2} - a'_{i1} a'_{12} . \qquad i \geq 2$$

Daarna de tweede rij aanvullen met:

$$a'_{2j} = \frac{1}{a'_{22}} (a_{2j} - a'_{21} a'_{1j}) \quad j > 2 \ ; \ b'_2 = \frac{1}{a'_{22}} (b_2 - a'_{21} b'_1) \ .$$

Op deze manier doorgaand berekent men de gehele matrix (7.3.3); uit (7.3.5) bepaalt men $x_n, x_{n-1}, \ldots, x_1$.

Als contrôle op de berekeningen kan men aan de matrix (7.3.2) toevoegen een kolom, waarvan de elementen gelijk zijn aan de som $c_i = \sum_{k=1}^{n} a_{ik} + b_i$. Behandelt men deze kolom c_i als b_i, dan vindt men een c'_i en een hierbij volgens (7.3.5) horende

$$y_i = c'_i - \sum_{k=i+1}^{n} a'_{ik} y_k \ .$$

$$\left.\begin{aligned} &\text{Nu geldt } c'_i = 1 + \sum_{k=i+1}^{n} a'_{ik} + b'_i \ , \\ &\text{en} \qquad y_i = 1 + x_i \end{aligned}\right\} \qquad (7.3.6)$$

(7.3.6) kan men als contrôle gebruiken. Ze behoeven natuurlijk slechts op afrondingsfouten na te kloppen.

7.4. Rechtvaardiging van de methode van Crout

Wij zullen uitgaande van de eliminatie van Gauss de formules van Crout afleiden.

$$\begin{matrix} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ \vdots \qquad\qquad \vdots \\ a_{n1}x_1 + \dots + a_{nn}x_n = b_n \end{matrix} \qquad (7.4.1)$$

Het stelsel na afloop van de Gauss-eliminatie zij:

$$\begin{matrix} x_1 + a'_{12}x_2 + \dots + a'_{1n}x_n = b'_1 \\ x_2 + \dots \quad \dots \quad = b'_2 \\ \dots \quad \dots \quad \vdots \\ x_n = b'_n \end{matrix} \qquad (7.4.2)$$

De k^e vergelijking van (7.4.2) is verkregen door van de k^e vergelijking van (7.4.1) een lineaire combinatie van de eerste k-1 vergelijkingen van (7.4.2) af te trekken. Hieruit volgt o.a. dat geldt:

$$b'_k = \alpha_0 b_k + \alpha_1 b'_{k-1} + \alpha_2 b'_{k-2} + \dots + \alpha_{n-1} b'_1 .$$

Dit leidt bij oplossing van b_k tot het stelsel:

$$\begin{matrix} a'_{11}b'_1 & & = b_1 \\ a'_{21}b'_1 + a'_{22}b'_2 & & = b_2 \\ \vdots & & \vdots \\ a'_{n1}b'_1 + & \dots & + a'_{nn}b'_n = b_n \end{matrix} \quad , \qquad (7.4.3)$$

waarbij als coëfficiënten gebruikt zijn a'_{ik}, $i \geq k$. Dit kan, omdat in (7.4.2) slechts voorkwamen a'_{ik} met $i < k$.

Laten wij nu nemen in plaats van (7.4.2) het stelsel

$$x_k + \sum_{j=1}^{n} \beta_{kj}\, x_j = b'_k \qquad (7.4.4)$$

met $\quad \beta_{kj} = 0 \quad$ als $\quad k \geq j$

$\qquad\qquad = a'_{kj} \quad$ als $\quad k < j$,

en in plaats van (7.4.3) het stelsel

$$\sum_{j=1}^{n} \alpha_{kj} b'_j = b_k \qquad (7.4.5)$$

met $\alpha_{kj} = a'_{kj}$ als $k \geqslant j$,

$\qquad\quad = 0$ als $k < j$.

Substitutie van (7.4.4) in (7.4.5) geeft

$$\sum_{j=1}^{n} \alpha_{kj}(x_j + \sum_{i=1}^{n} \beta_{ji} x_i) = b_k \qquad (7.4.6)$$

(7.4.6) is slechts een andere schrijfwijze voor (7.4.1):

$$\sum_{j=1}^{n} a_{kj}\, x_j = b_k \ . \qquad (7.4.7)$$

De coëfficiënten van x_j in (7.4.7) en (7.4.6) moeten dus aan elkaar gelijk zijn. Dit leidt tot:

$$a_{kj} = \alpha_{kj} + \sum_{i=1}^{n} \alpha_{ki} \beta_{ij} \qquad (7.4.8)$$

Uit (7.4.8) volgt met de definitie van α_{ki} en β_{ij}:

$$k \geqslant j \qquad a_{kj} = \alpha_{kj} + \sum_{i=1}^{j-1} \alpha_{ki} \beta_{ij} = a'_{kj} + \sum_{i=1}^{j-1} a'_{ki}\, a'_{ij} \ ,$$

hetgeen hetzelfde is als (7.3.4a).

Stelt men $k < j$, dan volgt uit (7.4.8)

$$a_{kj} = \sum_{i=1}^{k} \alpha_{ki} \beta_{ij} = \alpha'_{kk} \beta_{kj} + \sum_{i=1}^{k-1} \alpha_{ki} \beta_{ij} =$$

$$= a'_{kk}\, a'_{kj} + \sum_{i=1}^{k-1} a'_{ki}\, a'_{ij} \ ,$$

hetgeen hetzelfde is als (7.3.4b).

Tenslotte geeft (7.4.3)

$$a'_{kk}\, b'_k + \sum_{i=1}^{k-1} a_{ki}\, b'_i = b_k$$

en (7.4.2)

$$x_k + \sum_{i=k+1}^{n} a'_{ki}\, x_i = b'_k \quad ,$$

welke resp. hetzelfde zijn als (7.3.4c) en (7.3.5).

Hiermede is dus de juistheid van de formules van Crout aangetoond. De hulpmatrix van Crout is dus wat de elementen rechts van de hoofddiagonaal betreft gelijk aan die na afloop van de Gauss-eliminatie.

De juistheid van (7.3.6) kan men inzien door op te merken, dat bij verhoging van alle x_i met 1, b_i vermeerderd wordt met $\sum_{k=1}^{n} a_{ik}$ volgens (7.4.1) en dat b'_i volgens (7.4.2) vermeerderd wordt met $1 + \sum_{k=i+1}^{n} a'_{ik}$.
Omgekeerd heeft een kolom $b_i + \sum_{k=1}^{n} a_{ik}$ dus tot gevolg de bijbehorende kolommen b'_i en y_i volgens (7.3.6).

7.5. Toepassingen van de methode van Crout

Het inverteren van een matrix kan men met de methode van (7.3) doen door in plaats van de kolom b_i de eenheidsmatrix te nemen. Door ieder van zijn kolommen te behandelen alsof het de kolom b was vindt men n kolommen x_i, die de kolommen van de inverse matrix zijn.
Moet men een stelsel oplossen voor meer rechterleden, dan kan men dat zonder meer doen.

Heeft men een oplossing x_i gevonden, dan is deze behept met afrondingsfouten: men heeft in plaats van x_i gevonden $x_i + \delta x_i$. Substitueert men deze in (7.3.1), dan vindt men

$$\sum a_{ik}(x_k + \delta x_k) = b_i + \delta b_i \quad .$$

δx_k zou men kunnen bepalen uit $\sum a_{ik}\delta x_k = \delta b_i$. Hiertoe hoeft men dus slechts δb_i te bepalen, ze met een geschikte factor te vermenigvuldigen (ze zullen klein zijn) en daarna uit het stelsel δx_k op te lossen; van de hulpmatrix hoeft men alleen de laatste kolom opnieuw te berekenen.

7.6. Iteratieve Methoden

Laat de coëfficiëntenmatrix van een stelsel vergelijkingen zo zijn, dat het in absolute waarde grootste element van een rij op de hoofddiagonaal staat. Door verandering van de volgorde der vergelijkingen en onbekenden kan dit vaak bereikt worden. Het stelsel:

$$\sum_{k=1}^{n} a_{ik}x_k = b_i, \; i = 1(1)n \tag{7.6.1}$$

wordt geschreven als:

$$x_i = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{k=1 \\ k\neq i}}^{n} a_{ik}x_k\right), \; i = 1(1)n \tag{7.6.2}$$

Stel nu gegeven een schatting van de oplossing. Substitutie in het rechterlid van (7.6.2) geeft een nieuwe schatting; substitueert men deze weer enz., dan heeft men een iteratief procédé, dat van Gauss. Op de convergentie (zo die bestaat) zullen wij later ingaan.

Als men in het rechterlid van de eerste vergelijking van (7.6.2) een schatting substitueert krijgt men een nieuwe schatting voor x_1; deze kan men gebruiken bij de substitutie in de tweede vergelijking. Doet men dit steeds, d.w.z. maakt men steeds direct gebruik van nieuwe schattingen, dan itereert men volgens de methode Gauss-Seidel. Verwacht kan worden dat deze methode zo ze convergeert sneller convergeert dan die van Gauss.

De beide voorgaande methoden werken de vergelijkingen af in een vaste volgorde. Dit is niet het geval bij de relaxatie-methode. Hierbij schrijft men

$$R_i = \sum_{k=1}^{n} a_{ik}x_k - b_i \tag{7.6.3}$$

waarbij $R_i = 0$ is voor de oplossing x van (7.6.1). Substitueert men in (7.6.3) een schatting x, dan heeft R_i van nul verschillende waarden. Men zoekt nu die in absolute waarde grootste R_i en maakt hem nul door verandering van de x_k met de grootste coëfficiënt. Hierdoor veranderen alle residuen R_i. Van de nieuwe residuen zoekt men weer de grootste enz.

7.7. Voorbeelden

$$4x_1 + x_2 - 2x_3 = 8 \quad \text{heeft als oplossing} \quad x_1 = 1$$
$$2x_1 + 7x_2 - x_3 = 17 \qquad\qquad x_2 = 2$$
$$-3x_1 + x_2 - 8x_3 = 7 \qquad\qquad x_3 = -1$$

Wij schrijven nu

$$x_1 = \frac{1}{4}(8 - x_2 + 2x_3)$$
$$x_2 = \frac{1}{7}(17 - 2x_1 + x_3)$$
$$x_3 = -\frac{1}{8}(7 + 3x_1 - x_2)$$

en itereren met resp. Gauss en Gauss-Seidel, beginnend met $x_1 = x_2 = x_3 = 0$.

Gauss

x_1	0	2.0	.95	.91	.9975	1.0093
x_2	0	2.4	1.73	1.97	2.0229	2.0050
x_3	0	-.9	-1.32	-1.02	-.9700	-.9962

x_1	1.0006	.9991	.9999	1.0001	1.0000
x_2	1.9979	1.9994	2.0002	2.0001	2.0000
x_3	-1.0029	-1.0005	-.9997	-.9999	-1.0000

Gauss-Seidel

x_1	0	2.0000	.8393	1.0320	.9941	1.0011	.9994	1.0000
x_2	0	1.8571	1.9898	1.9993	2.0000	2.0020	2.0002	2.0000
x_3	0	-1.3929	-.9410	-1.0121	-.9978	-1.0002	-.9998	-1.0000

Relaxatie

$$R_1 = -8 + 4x_1 + x_2 - 2x_3$$
$$R_2 = -17 + 2x_1 + 7x_2 - x_3$$
$$R_3 = -7 - 3x_1 + x_2 - 8x_3$$

x_1	0	0	2	2	1
x_2	0	2	2	2	2
x_3	0	0	0	-1	-1

R_1	-8	-6	2	4	0
R_2	-17	-3	1	2	0
R_3	-7	-5	11	-3	0

7.8. Eigenwaarden van matrices

Onder eigenwaarden van een vierkante matrix verstaat men de waarden van λ waarvoor het stelsel $Ax = \lambda x$ een andere dan de nuloplossing (7.8.1) heeft. (7.8.1) kan men ook schrijven als:

$$(A - \lambda I)x = 0 \tag{7.8.2}$$

Opdat (7.8.2) een andere dan de nuloplossing heeft moet gelden

$$|A - \lambda I| = 0 \tag{7.8.3}$$

Als A een n x n-matrix is stelt (7.8.3) voor een n^e graads vergelijking in λ:

$$P_n(\lambda) = 0 \tag{7.8.4}$$

(7.8.4) heet de karakteristieke of seculair vergelijking; de wortels λ_i de eigenwaarden van A. Deze kunnen voor een reële matrix reëel of complex en enkel- of meervoudig zijn.

Is λ_i een enkelvoudig nulpunt, dan heeft

$$(A - \lambda_i I)x = 0 \tag{7.8.5}$$

één oplossingsvector x_i, die op een multiplicatieve factor na bepaald is: de bij λ_i behorende eigenvector x_i of ook wel eigenkolom x_i.

Is λ_i m-voudig en is de rang van $(A - \lambda_i I)$ gelijk aan n-m, dan heeft $(A - \lambda_i I)x = 0$ m onafhankelijke oplossingen x_i. Is de rang groter dan n-m, dan heeft men minder dan m oplossingen x_i bij λ_i. In onze beschouwingen zullen wij dit geval uitsluiten te meer, daar dit bij symmetrische matrices (A^T=A) het niet kan voorkomen.

Tot nu toe hebben wij slechts eigenkolommen beschouwd.

$$y^T(A - \lambda_i I) = 0 \tag{7.8.6}$$

heeft een van nul verschillende oplossing y_i^T, als λ_i een eigenwaarde van A is, de eigenrij behorende bij λ_i. Nu geldt niet, dat $y_i = x_i$. Opdat dit wel het geval zij, moet A^T=A zijn:

transponeer (7.8.6): $(A^T - \lambda_i I)y = 0$ dit is slechts hetzelfde stelsel als (7.8.5) als $A^T = A$.

Wij hebben bij een matrix A dus in het algemeen n eigenwaarden λ_i, n eigenkolommen x_i en n eigenrijen y_i.
Stel wij hebben x_i en y_k^T behorende resp. bij verschillende eigenwaarden λ_i en λ_k, dan $Ax_i = \lambda_i x_i$; $y_k^T A = \lambda_k y_k^T$.
Vermenigvuldig resp. met y_k^T en x_i:

$$y_k^T A x_i = \lambda_i y_k^T x_i ; \quad y_k^T A x_i = \lambda_k y_k^T x_i$$

De rechterleden zijn gelijk, $\lambda_i \neq \lambda_k$, dus $y_k^T x_i = 0$ (7.8.7)
Geldt $A^T = A$, dan geldt als $\lambda_i = \lambda_k$:

$Ax_i = \lambda_i k_i$; $Ax_k = \lambda_k x_k$; vermenigvuldig resp. met x_k^T en x_i^T:

$x_k^T A x_i = \lambda_i x_k^T x_i$; $x_i^T A x_i = \lambda_k x_k^T x_i$, dus daar $\lambda_i \neq \lambda_k$ geldt $x_k^T x_i = 0$.
Eigenvectoren van een symmetrische matrix behorende bij verschillende eigenwaarden zijn orthogonaal.
Heeft men twee verschillende eigenvectoren x_i en x_i^* behorende bij één λ_i van een symmetrische matrix A, dan is ook x_i
$x_i^{**} = x_i + \alpha x_i^*$ een eigenvector. Men kan de α zo kiezen, dat $x_i^T x_i^{**} = 0$ is:

$$0 = x_i^T x_i^{xx} = x_i^T x_i + \alpha x_i^T x_i^*, \text{ dus } \quad \alpha = -\frac{x_i^T x_i}{x_i^T x_i^*} \quad .$$

De noemer is $\neq 0$ wegens de veronderstelling dat x_i en x_i^* niet orthogonaal zijn. Wij hebben dus een nieuwe eigenvector x_i^{**} geconstrueerd, orthogonaal met x_i. Op deze wijze kan men alle bij één eigenwaarde behorende eigenvectoren orthogonaliseren:
Een symmetrische matrix heeft n orthogonale eigenvectoren.
N.B. Eigenkolommen en eigenrijen van symmetrische matrices zijn elkaars getransponeerde.
Verder geldt dat eigenwaarden van een reële symmetrische matrix reëel zijn. Stel λ_i was complex, dan:

$$Ax_i = \lambda_i x_i \quad \text{en} \quad A\bar{x}_i = \bar{\lambda}_i \bar{x}_i \quad \text{en}$$

$$\bar{x}_i^T A x_i = \lambda_i \bar{x}_i^T x_i \quad \text{en} \quad x_i^T A \bar{x}_i = \bar{\lambda}_i x_i^T \bar{x}_i \text{ of ook } \bar{x}_i^T A x_i = \bar{\lambda}_i \bar{x}_i^T x_i$$

$$(\lambda_i - \bar{\lambda}_i)\bar{x}_i^T x_i = 0 \quad \bar{x}_i^T x_i \neq 0, \text{ dus } \lambda_i = \bar{\lambda}_i \text{ en dus is}$$

λ_i reëel. Dit impliceert dat de eigenvectoren eveneens reëel zijn.

Zoals wij hebben gezien zijn eigenvectoren op een multiplicatieve constante na bepaald: deze bepalen wij door te stellen dat de lengte van een eigenvector gelijk 1 te stellen: $x_i^T x_i = 1$.

Wij zullen in het vervolg altijd deze z.g. genormeerde eigenvectoren gebruiken. Voor een niet-symmetrische matrix is de normeringsvoorwaarde $y_k^T x_k = 1$.

Beschouwt men de matrices

$$Y^T = \begin{pmatrix} y_1^{(1)} y_1^{(2)} \cdots y_1^{(n)} \\ \vdots \qquad \vdots \\ y_n^{(1)} y_n^{(2)} \cdots y_n^{(n)} \end{pmatrix} \text{ en } X = \begin{pmatrix} x_1^{(1)} \cdots x_n^{(1)} \\ \vdots \qquad \vdots \\ x_1^{(n)} \cdots x_n^{(n)} \end{pmatrix} \qquad (7.8.8)$$

dan geldt: $$Y^T A X = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \Lambda \qquad (7.8.9)$$

Y^T en X zijn de z.g. diagonaliserende matrices.

Voor symmetrische A vinden wij $X^T A X = \Lambda$ (7.8.10)

Nu ziet men gemakkelijk in met behulp van de normerings- en orthogonaliteitscondities, dat $XX^T = I$ dus $X^T = X^{-1}$, d.w.z. X is een orthogonale matrix.

Dus geldt: $$A = X\Lambda X^T \qquad (7.8.11)$$

En hieruit volgt: $$A^{-1} = X\Lambda^{-1}X^T \qquad (7.8.12)$$

Λ^{-1} is een diagonaal-matrix met op de diagonaal $\frac{1}{\lambda_i}$: de inverse A^{-1} kan men dus zeer eenvoudig bepalen als men X kent. Het bepalen van X is echter veel meer werk dan dat van A^{-1} volgens een van de andere methoden.

Voorbeeld

$$A = \begin{pmatrix} 261 & -70 & -115 \\ -70 & 36 & 10 \\ -115 & 10 & 101 \end{pmatrix}$$

$$A- \quad = \begin{vmatrix} 261-\lambda & -70 & -115 \\ -70 & 36-\lambda & 10 \\ -115 & 10 & 101-\lambda \end{vmatrix} = -(\lambda^3-21168\lambda^2+21168\lambda-112896)$$

waaruit dus $\lambda_1 = 336$, $\lambda_2 = 56$, $\lambda_3 = 6$.

Het stelsel behorende bij λ_1 wordt:

$$\begin{array}{rl} -75x_1 - 70x_2 - 115x_3 = 0 & \qquad x_1 = -4 \\ -70x_1 - 300x_2 + 10x_3 = 0 & \text{of} \quad x_2 = 1 \\ -115x_1 + 10x_2 - 235x_3 = 0 & \qquad x_3 = 2 \end{array}$$

$$\begin{array}{rl} \lambda_2: 205x_1 - 70x_2 - 115x_3 = 0 & \qquad x_1 = -1 \\ -70x_1 - 20x_2 + 10x_3 = 0 & \text{of} \quad x_2 = 2 \\ -115x_1 + 10x_2 + 45x_3 = 0 & \qquad x_3 = -3 \end{array}$$

$$\begin{array}{rl} \lambda_3: 255x_1 - 70x_2 - 115x_3 = 0 & \qquad x_1 = 1 \\ -70x_1 + 30x_2 + 10x_3 = 0 & \text{of} \quad x_2 = 2 \\ -115x_1 + 10x_2 + 95x_3 = 0 & \qquad x_3 = 1 \end{array}$$

Men vindt dus $\lambda_1 = 336$ en

$$\begin{array}{l} \lambda_2 = 56 \\ \lambda_3 = 6 \end{array} \qquad X = \begin{pmatrix} -4/\sqrt{21} & -1/\sqrt{14} & 1/\sqrt{6} \\ 1/\sqrt{21} & 2/\sqrt{14} & 2/\sqrt{6} \\ 2/\sqrt{21} & -3/\sqrt{14} & 1/\sqrt{6} \end{pmatrix}$$

waarbij de eigenvectoren op lengte 1 genormeerd zijn.

7.9. Iteratieve methoden ter bepaling van de grootste eigenwaarde en bijbehorende eigenvector van een symmetrische matrix.

Als eerste (veel gebruikte) methode behandelen wij de volgende: zij gegeven een vector $x^{(0)}$, vorm dan de rij $Ax^{(0)}, A^2x^{(0)}, \ldots$ $\ldots, A^kx^{(0)}$.

De bewering luidt, zo λ_1 de absoluut grootste eigenwaarde van A is, dat $A^kx^{(0)}$ voor grote k nadert tot x_1, de bij λ_1 behorende eigenvector en

$$\frac{|A^kx^{(0)}|}{|A^{k-1}x^{(0)}|} \text{ nadert tot } |\lambda_1| \; .$$

($A^kx^{(0)}$ is een vector; onder de absolute waarde $|x|$ van een vector x wordt verstaan $|x| = \sqrt{x^Tx}$.)

Dat de bewering juist is, zien wij als volgt in: laten gegeven zijn alle eigenwaarden λ_i met bijbehorende eigenvectoren x_i. Wij stellen x_i genormeerd, d.w.z. $|x_i| = 1$. Verder weten wij, dat $x_i^Tx_j = 0$ als $i \neq j$: immers A is symmetrisch. Tenslotte nemen wij aan, dat $|\lambda_i| > |\lambda_j|$ als $i < j$.
Wij stellen nu

$$x^{(0)} = \alpha_1x_1 + \alpha_2x_2 + \ldots + \alpha_nx_n \qquad (7.9.1)$$

en veronderstellen dat $\alpha_1 \neq 0$. Dit betekent x niet loodrecht op x_1. ($x_1^Tx^{(0)} = \alpha_1x_1^Tx_1 = \alpha$)

Voor-vermenigvuldiging van (7.9.1) met A geeft:

$$x^{(1)} = Ax^{(0)} = \alpha_1Ax_1 + \alpha_2Ax_2 + \ldots + \alpha_nAx_n =$$
$$\alpha_1\lambda_1x_1 + \alpha_2\lambda_2x_2 + \ldots + \alpha_n\lambda_nx_n$$

en algemeen $x^{(k)} = A^kx^{(0)} = \alpha_1\lambda_1^kx_1 + \alpha_2\lambda_2^kx_2 + \ldots + \alpha_n\lambda_n^kx_n$

$$(7.9.2)$$

(7.9.2) is een vectorvergelijking evenals (7.9.1) en geldt voor ieder element van $x^{(k)}$, bijv. voor het j^e element geldt:

$$x_j^{(k)} = \alpha_1\lambda_1^kx_{1_j} + \alpha_2\lambda_2^kx_{2_j} + \ldots + \alpha_n\lambda_n^kx_{n_j} \qquad (7.9.3)$$

Nu gebruiken wij dat $|\lambda_1| > |\lambda_i|$, $i \neq 1$ en zien dat bij grote k de eerste term gaat overheersen:

$$x_j^{(k)} = \alpha_1\lambda_1^k \left\{x_{1_j} + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k x_{2_j} + \frac{\alpha_3}{\alpha_1}\left(\frac{\lambda_3}{\lambda_1}\right)^k x_{3_j} + \ldots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k x_{n_j}\right\} \tag{7.9.4}$$

Wij zien dat afgezien van de factor $\alpha_1\lambda_1^k$ die niet van j afhangt $x_j^{(k)}$ nadert tot x_{1_j}, het j^e element van de eerste eigenvector en wel met een fout die iedere keer met $\frac{\lambda_2}{\lambda_1}$ vermenigvuldigd wordt.
De factor $\alpha_1\lambda_1^k$ kunnen wij door normering kwijtraken.
Voor het bepalen van $|\lambda_1|$ berekenen wij $|x^{(k)}|$:

$$x^{(k)T}x^{(k)} = (\alpha_1\lambda_1^k x_1^T + \alpha_2\lambda_2^k x_2^T + \ldots + \alpha_n\lambda_n^k x_n^T)(\alpha_1\lambda_1^k x_1 + \alpha_2\lambda_2^k x_2 + \ldots \alpha_n\lambda_n^k x_n) =$$

$$= \alpha_1^2\lambda_1^{2k} + \alpha_2^2\lambda_2^{2k} + \ldots + \alpha_n^2\lambda_n^{2k} \tag{7.9.5}$$

Dus $$|x^{(k)}| = \sqrt{\alpha_1^2\lambda_1^{2k} + \alpha_2^2\lambda_2^{2k} + \ldots + \alpha_n^2\lambda_n^{2k}} =$$

$$= |\alpha_1\lambda_1^k| \sqrt{1 + \left(\frac{\alpha_2}{\alpha_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k} + \ldots} = |\alpha_1\lambda_1^k|\left\{1 + \tfrac{1}{2}\left(\frac{\alpha_2}{\alpha_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k} + \ldots\right\} \tag{7.9.6}$$

En
$$\frac{|x^{(k)}|}{|x^{(k-1)}|} = \frac{\alpha_1\lambda_1^k}{\alpha_1\lambda_1^{(k-1)}} \; \frac{\left\{1+\frac{1}{2}\left(\frac{\alpha_2}{\alpha_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}+\ldots\right\}}{\left\{1+\frac{1}{2}\left(\frac{\alpha_2}{\alpha_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k-2}+\ldots\right\}} =$$

$$= |\lambda_1| \left\{1 - \tfrac{1}{2}\left(\frac{\alpha_2}{\alpha_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k-2}\left(1 - \left(\frac{\lambda_2}{\lambda_1}\right)^2\right) + \ldots\right\} \tag{7.9.7}$$

Uit (7.9.7) volgt dus dat $\frac{|x^{(k)}|}{|x^{(k-1)}|}$ nadert tot $|\lambda_1|$ met een fout die bij iedere iteratie met $\left(\frac{\lambda_2}{\lambda_1}\right)^2$ vermenigvuldigd wordt.
Bij de iteratie convergeert de eigenwaarde dus sneller dan de eigenvector. Dit geldt alleen voor symmetrische matrices, zoals later zal blijken.
Het teken voor λ_1 vindt men door voor voldoende grote k, bijv. de laatste en zijn voorganger twee overeenkomstige elementen van $x^{(k)}$ en $x^{(k-1)}$ te vergelijken: zijn de tekens gelijk, dan is $\lambda_1 > 0$, zijn ze ongelijk, dan $\lambda_1 < 0$.

Bij bovenstaande methode zijn een aantal opmerkingen te maken:

1. De iteratie convergeert tot λ_1 en x_1 als $\alpha_1 \neq 0$. Het kan zijn dat men een bijzonder slechte beginschatting $x^{(0)}$ gekozen heeft waarvoor α_1 ongeveer nul is. Door afrondingsfouten zal er toch een component in de x_1-richting komen: het duurt alleen langer tot men λ_1 en x_1 gevonden heeft.

2. Wat betreft de grootte van de optredende elementen van vectoren $x^{(k)}$ het volgende: als $|\lambda_1| > 1$ dan zullen volgens (7.9.2) de elementen van $x^{(k)}$ met toenemende k groot worden; is $|\lambda_1| < 1$ dan worden ze juist zeer klein; dit is onaangenaam en kan eenvoudig verholpen worden. Wij itereren als volgt:

$$\begin{aligned} y^{(k+1)} &= Ax^{(k)} \\ x^{(k+1)} &= \frac{y^{(k+1)}}{|y^{(k+1)}|} \end{aligned} \qquad (7.9.8)$$

waarbij als $|x^{(0)}| = 1$ gekozen wordt. $|x^{(k)}|$ blijft dan 1. In de zin van de eerst gegeven iteratie zijn $y^{(k+1)}$ en $x^{(k)}$ opvolgende iteratie-resultaten. Daarmee gaat (7.9.6) over in

$$\begin{aligned} &\frac{|y^{(k+1)}|}{|x^{(k)}|} \approx |\lambda_1| \quad \text{naar} \quad |x^{(k)}| = 1, \text{ dus} \\ &|y^{(k+1)}| \approx |\lambda_1| \end{aligned} \qquad (7.9.9)$$

m.a.w. de norm van $y^{(k)}$ nadert tot $|\lambda_1|$.

3. Aangenomen was $|\lambda_i| > |\lambda_j|$ $i < j$. Als $\lambda_1 = \lambda_2$, dan komen er moeilijkheden: in (7.9.2) blijven nu de eerste twee termen van dezelfde orde. Wij vinden dan een eigenvector die bestaat uit een mengsel van x_1 en x_2. Echter als $\lambda_1 = \lambda_2$, dan is ook $x = \alpha x_1 + \beta x_2$ een eigenvector. Het gelijk zijn van λ_1 en λ_2 geeft dus schijnmoeilijkheden. Is $\lambda_1 = -\lambda_2$, dan geeft dit ook geen moeilijkheden, mits men de iteraties om de andere bekijkt. Dan itereert men als het ware met A^2, die twee gelijke eigenwaarden λ_1^2 en λ_2^2 en dezelfde eigenvectoren als A heeft. Wij hebben dus het vorige geval terug. Erger is het als $|\lambda_1|$ ongeveer gelijk $|\lambda_2|$ is. Dan is $x^{(k)}$, tenzij k extreem groot is,

nog steeds een mengsel van x_1 en x_2: nu is $x = \alpha x_1 + \beta x_2$ geen eigenvector. Dit probleem wordt later behandeld.

Voorbeeld

$$A = \begin{pmatrix} 261 & -70 & -115 \\ -70 & 36 & 10 \\ -115 & 10 & 101 \end{pmatrix}$$

Laat zijn

$$x^{(0)} = \begin{pmatrix} .5774 \\ .5774 \\ .5774 \end{pmatrix},$$

dan

$$y_1^{(1)} = 261 \times .5774 - 70 \times .5774 - 115 \times .5774 = 43.88$$

$$y_2^{(1)} = -70 \times .5774 + 36 \times .5774 + 10 \times .5774 = -13.86$$

$$y_3^{(1)} = -115 \times +5774 + 10 \times .5774 + 101 \times .5774 = -2.31$$

$$y_1 = \sqrt{43.88^2 + 13.86^2 + 2.31^2} = 46.07$$

$$x_1^{(1)} = \frac{y_1^{(1)}}{|y_1|} = .9524$$

$$x_2^{(1)} = -.3008$$

$$x_3^{(1)} = -.0501$$

Hiermede verder itererend vinden wij

	$y^{(0)}$	$x^{(0)}$	$y^{(1)}$	$x^{(1)}$	$y^{(2)}$	$x^{(2)}$
	1.0000	.5774	43.88	.9524	275.39	.8900
$\|y\|^{(k)}$	1.0000	.5774	-13.86	-.3008	- 78.00	-.2521
	1.0000	.5774	- 2.31	-.0501	-117.59	-.3800
			46.07		309.44	

	$y^{(3)}$	$x^{(3)}$	$y^{(4)}$	$x^{(4)}$	$y^{(5)}$	$x^{(5)}$
	293.64	.8759	293.44	.8734	293.31	.8730
$\|y\|^{(k)}$	- 75.18	-.2242	-73.66	-.2192	-73.38	-.2184
	-143.25	-.4273	-146.13	-.4349	-146.56	-.4362
	335.26		335.99		336.00	

	$y^{(6)}$	$x^{(6)}$	$y^{(7)}$	$x^{(7)}$
	293.30	.8729	293.29	.8729
$\|y\|^{(k)}$	-73.33	-.2182	-73.32	-.2182
	-146.64	-.4364	-146.64	-.4364
	336.01		336.00	

Dus

$$\lambda_1 = +\ 336.00 \qquad x_1 = \begin{pmatrix} .8729 \\ -.2182 \\ .4364 \end{pmatrix} .$$

7.10. Hogere eigenwaarden van symmetrische matrices

In de vorige paragraaf hebben wij gezien, dat de daar gegeven methode slechts de eerste eigenwaarde en eigenvector levert, wat de beginschatting ook zij.
Om een volgende te vinden staan ons twee mogelijkheden open:

1. $\alpha_1 = 0$ maken en zorgen dat $\alpha_1 = 0$ blijft.
2. $\lambda_1 = 0$ maken.

In de eerste opzet slaagt men door met willekeurige x te nemen

$$x^{(0)} = x + \alpha x_1 \text{ zo, dat } x_1^T x^{(0)} = 0, \text{ dus } \alpha = -x_1^T x \text{ of} \qquad (7.10.1)$$
$$x^{(0)} = x - x_1^T x . x_1$$

Nu is dus $\alpha_1 = x_1^T x^{(0)} = x_1^T x - x_1^T x . x_1^T x_1 = 0$

Men moet erop bedacht zijn, dat de afrondingsfouten ervoor zorgen dat α_1 niet nul blijft. Men houdt α_1 door iedere keer de component in de x_1-richting uit te schakelen zo goed mogelijk nul.
Heeft men een hogere dan de tweede eigenwaarde te bepalen, bijv. de k^e, dan moet men steeds orthogonaliseren t.o.v. de eerste k-1 eigenvectoren: men moet dan k-1 α's bepalen.

Voorbeeld

$$A = \begin{pmatrix} 261 & -70 & -115 \\ -70 & 36 & 10 \\ -115 & 10 & 101 \end{pmatrix} \qquad x_1 = \begin{pmatrix} .8729 \\ -.2182 \\ -.4364 \end{pmatrix}$$

Wij gebruiken:
$$\left.\begin{aligned} y^{(x)} &= Ax^{(k-1)} \\ z^{(k)} &= \frac{y^{(k)}}{|y^{(k)}|} \\ x^{(k+1)} &= z^{(k)} - x_1^T z^{(k)} x_1 \end{aligned}\right\} \qquad (7.10.2)$$

De tweede vergelijking geeft de normalisering, de derde de orthogonalisering t.o.v. x_1

	$y^{(0)}$	$z^{(0)}$	$x^{(0)}$	$y^{(1)}$	$z^{(1)}$	$x^{(1)}$
	1.000	.5774	.4674	6.9224	.3806	.3809
	1.000	.5774	.6049	- 4.6176	-.2539	-.2540
	1.000	.5774	.6324	16.1704	.8892	.8890
$y^{(k)}$				18.1858		
$-x_1^T z^{(k)}$					.0004	

	$y^{(2)}$	$z^{(2)}$	$x^{(2)}$	$y^{(3)}$	$z^{(3)}$	$x^{(3)}$
	14.9599	.2809	.2813	15.0173	.2683	.2687
	-26.9170	-.5055	-.5056	-29.7366	-.5313	-.5314
	43.4455	.8158	.8156	44.9701	.8035	.8033
$y^{(k)}$	53.2526			55.9651		
$-x_1^T z^{(k)}$		.0005			.0005	

	$y^{(4)}$	$z^{(4)}$	$x^{(4)}$	$y^{(5)}$	$z^{(5)}$	$x^{(5)}$
	14.9492	.2670	.2674	14.9554	.2671	.2673
	-29.9064	-.5341	-.5342	-29.9292	-.5344	-.5344
	44.9188	.8020	.8020	44.9090	.8019	.8018
$y^{(k)}$	55.9962			56.0021		
$-x_1^T z^{(k)}$					.0002	

	$y^{(6)}$	$z^{(6)}$	$x^{(6)}$
	14,9663	.2673	.2673
	29.9314	-.5345	-.5345
	44.8983	.8018	.8018
$y^{(k)}$	55.9976		
$-x_1^T z^{(k)}$		.0000	

Dus

$$_2 = 55.9976 \qquad x_2 = \begin{pmatrix} .2673 \\ -.5345 \\ .8018 \end{pmatrix} .$$

Het nadeel van deze methode is, dat het werk verbonden aan de orthogonalisatie evenredig is met het aantal iteraties.
Dit heeft men niet bij de deflatie volgens Hotelling.

Bij deze methode vervangt men de matrix A door een andere A_1, die dezelfde eigenvectoren heeft, maar waarvan de eigenwaarde behorende bij x_1 nul is:

$$A_1 = A - \lambda_1 x_1 x_1^T \tag{7.10.3}$$

$x_1 x_1^T$ is een symmetrische matrix, dus A_1 ook. Beschouw:

$$A_1 x_k = A x_k - \lambda_1 x_1 x_1^T x_k$$

waarbij x_k een eigenvector is; als $k \neq 1$ is dan vinden wij

$$A_1 x_k = A x_k = \lambda_k x_k,$$

dus λ_k is een eigenwaarde van A_1 en x_k de bijbehorende vector.
Is $k = 1$, dan:

$$A_1 x_1 = A x_1 - \lambda_1 x_1 x_1^T x_1 = \lambda_1 x_1 - \lambda_1 x_1 = 0 .$$

Bij de eigenvector x_1 behoort dus de eigenwaarde 0, m.a.w. als wij A_1 gaan itereren, volgens par. 7.9, dan vinden wij λ_2, x_2.

Voorbeeld

$$A = \begin{pmatrix} 261 & -70 & -115 \\ -70 & 36 & 10 \\ -115 & 10 & 101 \end{pmatrix} \qquad x_1 = \begin{pmatrix} .8729 \\ -.2182 \\ -.4364 \end{pmatrix} \qquad \lambda_1 = 335.00$$

Men vindt voor de eerste kolom van $x_1 x_1^T$:

$$\begin{pmatrix} .8729 \times .8729 \\ -.2182 \times .8729 \\ -.4364 \times .8729 \end{pmatrix}$$

en vervolgens

$$\lambda_1 x_1 x_1^T = \begin{pmatrix} 256.02 & -64.00 & -127.99 \\ -64.00 & 16.00 & 31.99 \\ -127.99 & 31.99 & 63.99 \end{pmatrix}$$

en

$$A_1 = \begin{pmatrix} 4.98 & -6.00 & 12.99 \\ -6.00 & 20.00 & -21.99 \\ 12.99 & -21.99 & 37.01 \end{pmatrix}$$

Toepassing van de iteratie op A_1 van par. 7.9 geeft:

	$y^{(0)}$	$x^{(0)}$	$y^{(1)}$	$x^{(1)}$	$y^{(2)}$	$x^{(2)}$	
	1.0000	.5774	6.91	.3801	14.97	.2811	
	1.0000	.5774	- 4.61	-.2537	-26.91	-.5054	
	1.0000	.5774	16.17	.8895	43.44	.8158	
$y^{(k)}$			18.18		53.25		
	$y^{(3)}$	$x^{(3)}$	$y^{(4)}$	$x^{(4)}$	$y^{(5)}$	$x^{(5)}$	$y^{(6)}$
	15.03	.2686	14.96	.2672	14.95	.2670	14.95
	-29.73	-.5313	-29.90	-.5340	-29.92	-.5344	-29.92
	44.96	.8034	44.91	.8021	44.90	.8019	44.90
	55.96		55.99		55.99		

Dus $\lambda_2 = 55.99 \qquad x_2 = \begin{pmatrix} .2670 \\ -.5344 \\ .8019 \end{pmatrix}$

Vervolgens vormen wij $A_2 = A_1 - \lambda_2 x_2 x_2^T$

$$A_2 = \begin{pmatrix} .99 & 1.99 & 1.00 \\ 1.99 & 4.01 & 2.00 \\ 1.00 & 2.00 & 1.01 \end{pmatrix}$$

Iteratie op A_2 geeft:

$y^{(0)}$	$x^{(0)}$	$y^{(1)}$	$x^{(1)}$	$y^{(2)}$	$x^{(2)}$	$y^{(2)}$
1.0000	.5774	2.30	.4064	2.44	.4067	2.44
1.0000	.5774	4.62	.8163	4.90	.8167	4.90
1.0000	.5774	2.32	.4099	2.45	.4083	2.45
		5.66		6.00		

en dus $\lambda_3 = 6.00 \qquad x_3 = \begin{pmatrix} .4067 \\ .8167 \\ .4083 \end{pmatrix}.$

Nogmaals deflatie geeft een matrix met alle eigenwaarden nul: de nulmatrix. Numeriek is dit natuurlijk wegens afrondingsfouten niet geheel het geval:

$$\cdot\ A_3 = \begin{pmatrix} .00 & .00 & .00 \\ .00 & .01 & .00 \\ .00 & .00 & ;01 \end{pmatrix}$$

Wij hebben dus gevonden de eigenvectorenmatrix

$$X = \begin{pmatrix} .8729 & .2670 & .4067 \\ -.2182 & -.5344 & .8167 \\ -.4363 & .8019 & .4083 \end{pmatrix} \quad \text{en} \quad \Lambda = \begin{pmatrix} 336.00 & 0 & 0 \\ 0 & 56.00 & 0 \\ 0 & 0 & 6.00 \end{pmatrix}$$

terwijl de oplossing had moeten zijn (pag. 40):

$$X = \begin{pmatrix} .8729 & .2673 & .4082 \\ -.2182 & -.5345 & .8165 \\ -.4364 & .8018 & .4082 \end{pmatrix}$$

7.11. Bijna gelijke eigenwaarden; methode van de vierkantsvergelijking

Wij hebben gezien dat bijna gelijke eigenwaarden bij de iteratie langzame convergentie tot gevolg hebben; de fout in de eigenvector werd immers bij iedere stap vermenigvuldigd met $\left|\frac{\lambda_i}{\lambda_{i-1}}\right|$. Wij nemen nu aan dat λ_1 en λ_2 ongeveer gelijke modulus hebben. Wij vinden, als wij zonder tussentijdse normering itereren:

$$\begin{aligned} x^{(k-1)} &= \alpha_1\lambda_1^{k-1}x_1 + \alpha_2\lambda_2^{k-1}x_2 + \ldots + \alpha_n\lambda_n^{k-1}x_n \\ x^{(k)} &= \alpha_1\lambda_1^{k}x_1 + \alpha_2\lambda_2^{k}x_2 + \ldots + \alpha_n\lambda_n^{k}x_n \qquad (7.11.1) \\ x^{(k+1)} &= \alpha_1\lambda_1^{k+1}x_1 + \alpha_2\lambda_2^{k+1}x_2 + \ldots + \alpha_n\lambda_n^{k+1}x_n \end{aligned}$$

Wij denken nu de k zo hoog opgevoerd, dat alle hogere eigenwaarden λ_3 t/m λ_n geen rol meer spelen.
Wij hebben dan

$$\begin{aligned} x^{(k-1)} &\approx \alpha_1\lambda_1^{k-1}x_1 + \alpha_2\lambda_2^{k-1}x_2 \\ x^{(k)} &\approx \alpha_1\lambda_1^{k}x_1 + \alpha_2\lambda_2^{k}x_2 \qquad (7.11.2) \\ x^{(k+1)} &\approx \alpha_1\lambda_1^{k+1}x_1 + \alpha_2\lambda_2^{k+1}x_2 \quad . \end{aligned}$$

Hierbij is aangenomen, dat niet tussentijds genormeerd is. (7.11.2) zullen wij in de veronderstelling dat gelijktekens gelden, oplossen. Stel daartoe: λ_1 en λ_2 voldoen aan een 2^e-graadsvergelijking:

$$\lambda^2 + \beta\lambda + \gamma = 0 \tag{7.11.3}$$

Vermenigvuldig de 1e vergelijking van (7.11.2) met γ, de 2e met β en de 3e met 1 en tel op. Wegens

$$\lambda_{1,2}^2 + \beta\lambda_{1,2} + \gamma = 0 \tag{7.11.4}$$

wordt het rechterlid nul:

$$x^{(k+1)} + \beta x^{(k)} + \gamma x^{(k-1)} = 0 \tag{7.11.5}$$

(7.11.5) is weer een vectorvergelijking die geldt (bij benadering omdat in (7.11.2) geen gelijktekens staan) voor ieder element $x_j^{(k)}$. Wij nemen twee elementen j en l; kies bijv. die die in de grootste precisie bekend zijn.
Dus

$$\left.\begin{aligned} x_j^{(k+1)} + \beta x_j^{(k)} + \gamma x_j^{(k-1)} &= 0 \\ x_l^{(k+1)} + \beta x_l^{(k)} + \gamma x_l^{(k-1)} &= 0 \end{aligned}\right\} \tag{7.11.6}$$

Uit (7.11.6) lossen wij op β en γ en vervolgens uit (7.11.3) λ_1 en λ_2. Daarna lossen wij $\alpha_1 x_{1_i}$ en $\alpha_2 x_{2_i}$ op uit:

$$\begin{aligned} x_i^{(k)} &= \alpha_1 \lambda_1{}^k x_{1_i} + \alpha_2 \lambda_2{}^k x_{2_i} \\ x_i^{(k+1)} &= \alpha_1 \lambda_1{}^{k+1} x_{1_i} + \alpha_2 \lambda_2{}^{k+1} x_{2_i} \end{aligned} \tag{7.11.7}$$

Dit geeft:

$$x_{1_i} = \frac{\lambda_2^{k+1} x_i^k - \lambda_2^k x_i^{k+1}}{\lambda_1^k \lambda_2^k (\lambda_2 - \lambda_1)} = \frac{x_i^{k+1} - \lambda_2 x_i^k}{\lambda_1^k (\lambda_1 - \lambda_2)}$$

$$x_{2_i} = \frac{\lambda_1^k x_i^{k+1} - \lambda_1^{k+1} x_i^k}{\lambda_1^k \lambda_2^k (\lambda_2 - \lambda_1)} = \frac{-x_i^{k+1} + \lambda_1 x_i^k}{\lambda_2^k (\lambda_1 - \lambda_2)}$$

waarbij de constante in de noemer er niet toe doet: wij normeren hem toch weg:

$$\left.\begin{aligned} x_{1_i} &= x_i^{k+1} - \lambda_2 x_i^{k} \\ x_{2_i} &= -x_i^{k} + \lambda_1 x_i^{k+1} \end{aligned}\right\} \quad i = 1(1)4 \qquad (7.11.8)$$

Als wij dit normeren hebben wij een schatting voor x_1 en x_2, die exact was als de in (7.11.1) weggelaten termen nul waren. Hiermee itereren wij verder. Zijn λ_1 en x_1 constant, dan wordt deflatie toegepast. Met λ_2 heeft men dan geen moeite meer.

Itereert men met tussentijdse normering, dan kan men als drie schattingsvectoren gebruiken

$$x^{(k)},\ y^{(k+1)},\ |y^{(k+1)}|.y^{(k+2)} \qquad (7.11.9)$$

zoals men direct inziet.

<u>Voorbeeld</u>

$$A = \begin{pmatrix} -.7057 & +.0114 & +.5829 \\ +.0114 & +.1771 & -.5657 \\ +.5829 & -.5657 & +.4486 \end{pmatrix}$$

$y^{(0)}$	$x^{(0)}$	$y^{(1)}$	$x^{(1)}$	$y^{(2)}$	$x^{(2)}$	$y^{(3)}$	$x^{(3)}$
1.0000	.5774	-.0643	-.1827	.5673	.5786	-.0661	-.0650
1.0000	.5774	-.2177	.6185	-.5439	-.5547	-.4299	-.4227
1.0000	.5774	.2690	.7642	.5862	.5979	.9193	.9039
		.3520		.9805		1.0170	

$y^{(4)}$	$x^{(4)}$	$y^{(5)}$	$x^{(5)}$	$y^{(6)}$	$x^{(6)}$	$y^{(7)}$
.5679	.5582	-.0529	-.0520	.5584	.5487	-.0411
-.5869	-.5769	-.4331	-.4257	-.5870	-.5768	-.4382
.6067	.5963	.9192	.9034	-.6158	.6051	.9176
1.0174		1.0175		1.0176		1.0177

De convergentie naar x_1 is niet snel, daarom vierkantsvergelijking opstellen. De iteratievectoren die wij gebruiken, zijn $x^{(5)}$, $y^{(6)}$ en $|y|^{(6)}.y^{(7)}$:

$x^{(5)}$	$y^{(6)}$	$\lvert y^{(6)}\rvert . y^{(7)}$
-.0520	.5584	-.0418
-.4257	-.5870	-.4459
.9034	.6158	.9337

Gebruik makend van (7.11.6) met j = 2 en l = 3 vinden wij

$$.5870\beta + .4257\gamma = -.4459$$
$$.6158\beta + .5034\gamma = -.9337$$

en

$$\beta = -.0196$$
$$\gamma = -1.0199 \ .$$

Vervolgens uit (7.11.3)

$$\lambda_1 = 1.0198$$
$$\lambda_2 = -1.0002$$

en tenslotte uit (7.11.8) de schattingen voor x_1 en x_2:

x_1^{**}	x_2^{**}
.5167	+.6113
-1.0330	-.1528
1.5496	-.3057

en na normering:

x_1^{*}	x_2^{*}
.2673	.8729
-.5345	-.2182
.8018	-.4365

Iteratie van x_1^* geeft $\lambda_1 = 1.0200$ en $x_1^* = x_1$. Men kan nu defleren en x_2^* itereren enz.

7.12. Niet symmetrische matrices

Zoals wij in 7.8 gezien hebben heeft men bij niet-symmetrische matrices te maken met n eigenkolommen en n eigenrijen, die resp. voldoen aan:

$$Ax = \lambda x \quad \text{en} \quad y^T A = \lambda y^T \tag{7.12.1}$$

waarbij de n eigenwaarden van kolom- en rij-vectoren gelijk zijn. Verder weten wij dat als $\lambda_i \neq \lambda_k$ is, dat

$$y_k^T x_i = 0 \tag{7.12.2}$$

Wij kunnen y_k^T en x_i zo normeren, dat

$$y_k^T x_k = 1 \quad \text{voor} \quad k = 1(1)n \ . \tag{7.12.3}$$

Vormt men de matrix X, die alle eigenkolommen tot kolommen heeft, dan geldt als Λ de diagonaalmatrix van de eigenwaarden is:

$$AX = X\Lambda \ . \tag{7.12.4}$$

Vormt men Y^T die tot rijen de rijen y_k^T heeft, dan

$$Y^T A = \Lambda Y^T \tag{7.12.5}$$

Nu volgt uit (7.12.2) en (7.12.3) dat $Y^T X = I$, dus

$$Y^T = X^{-1} \tag{7.12.6}$$

Achtervermenigvuldiging van (7.12.4) met Y^T geeft dus:

$$A = X\Lambda Y^T = X\Lambda X^{-1} \tag{7.12.7}$$

in tegenstelling tot (7.8.11) waar $X^T = X^{-1}$ was.

Om de eigenkolommen en -rijen te bepalen kan men weer itereren, bijv. voor de kolommen met A, voor de rijen met A^T.
Wat betreft de convergentie is er een verschil: bij de afleiding van (7.9.7) is gebruikt (7.9.5); nu komt:

$$x^{(k)^T} x^{(k)} = (\alpha_1\lambda_1^k x_1^T + \alpha_2\lambda_2^k x_2^T + \dots + \alpha_n\lambda_n^k x_n^T)(\alpha_1\lambda_1^k x_1 + \dots + \alpha_n\lambda_n^k x_n) =$$

$$= \alpha_1^2\lambda_1^{2k} x_1^T x_1 + 2\alpha_1\lambda_1^k(\alpha_2\lambda_2^k x_1^T x_2 + \dots + \alpha_n\lambda_n^k x_1^T x_n) + \text{termen zonder } \lambda_1 \tag{7.12.8}$$

Wij vinden dus:

$$|x^{(k)}| = |\alpha_1\lambda_1^k|\,|x_1|\sqrt{1+\frac{2\{\alpha_1\lambda_1^k\}^{-1}}{x_1^Tx_1}(\alpha_2\lambda_2^kx_1^Tx_2+\ \dots\ +\alpha_n\lambda_n^kx_1^Tx_n) + \dots} \qquad (7.12.9)$$

$$= |\alpha_1\lambda_1^k|\,|x_1|(1 + \frac{\alpha_2}{\alpha_1}(\frac{\lambda_2}{\lambda_1})^k\ \frac{x_1^Tx_2}{x_1^Tx_1} + \dots)$$

en

$$\frac{|x^{(k)}|}{|x^{(k-1)}|} = |\lambda_1|\ (1 - \frac{\alpha_2}{\alpha_1}\ (\frac{\lambda_2}{\lambda_1})^{k-1}\ \frac{x_1^Tx_2}{x_1^Tx_1} + \dots) \qquad (7.12.10)$$

De fout in λ_1 wordt dus nu steeds met $\frac{\lambda_2}{\lambda_1}$ vermenigvuldigd i.p.v. met $(\frac{\lambda_2}{\lambda_1})^2$ zoals bij symmetrische matrices het geval was.
Wat betreft de deflatie verandert er iets ten gevolge van het feit dat $x_k^Tx_i \neq 0$ is. Wij defleren nu een kolom en rij tegelijkertijd:

$$A_1 = A - \lambda_1x_1y_1^T \quad \text{als} \quad y_i^Tx_i = 1 \qquad (7.12.11)$$

De matrix A_1 bevat niet meer de eigenwaarde λ_1:

$$A_1x_1 = Ax_1 - \lambda_1x_1y_1^Tx_1 = \lambda_1x_1 - \lambda_1x_1 = 0$$

$$y_1^TA_1 = y_1^TA - \lambda_1y_1^Tx_1y_1^T = \lambda_1y_1^T - \lambda_1\ {}_1^T = 0$$

$$(7.12.12)$$

en alle andere λ_i zijn nog aanwezig:

$$A_1x_i = Ax_i - \lambda_1x_1y_1^Tx_i = A_ix_i$$

Als $y_i^Tx_i \neq 1$ is, dan moet men nemen:

$$A_1 = A - \lambda_1\ \frac{x_1y_1^T}{y_1^Tx_1} \qquad (7.12.13)$$

Om te kunnen defleren moet men dus rij en kolom kennen; dan kan men weer een rij en kolom itereren, enz. De twee waarden van λ_i die men vindt zullen niet precies gelijk zijn; bij de deflatie neme men een gemiddelde, bijv.:

$$\lambda_1 = \frac{\lambda_1^{(1)}+\lambda_1^{(2)}}{2}\ .$$

Bij niet symmetrische matrices kunnen complexe eigenwaarden, dus ook complexe eigenvectoren optreden. Deze kan men alleen met de vierkantsvergelijking bepalen: de moduli zijn immers gelijk. Men vindt dan dat de vierkantsvergelijking twee toegevoegd complexe wortels heeft. Men vindt van schattingen voor de twee (toegevoegd complexe) eigenvectoren, die men niet verder kan itereren: men moet met de oude schatting nog een iteratie uitvoeren en weer de vierkantsvergelijking oplossen, en kijken of de eigenvector veranderd is. Dit laatste is niet eenvoudig: de eigenvector is niet genormeerd. Men kan een complexe eigenvector normeren door de som der kwadraten van reële en imaginaire delen 1 te maken: dus door te vormen $\frac{x}{\sqrt{\bar{x}^T x}}$ waarbij $\bar{x}$ de toegevoegd complexe van x beduidt.

Normeert men de complexe vectoren aldus, dan zijn zij op een complexe factor met absolute waarde 1 na bepaald. Men kan ze dus nog moeilijk vergelijken.

Beter kan men als x_i een schatting voor de eigenvector is, kijken of

$$Ax_i - \lambda_i x_i$$

in de gewenste precisie nul is.

Bij de deflatie van complexe eigenwaarden moet men om de matrix reëel te houden λ_1 en $\bar{\lambda}_1$ tegelijkertijd defleren:

$$A_2 = A - \lambda_1 \frac{x_1 y_1^T}{y_1^T x_1} - \bar{\lambda}_1 \frac{\bar{x}_1 \bar{y}_1^T}{\bar{y}_1^T \bar{x}_1} = A - 2 \text{ Re } \lambda_1 \frac{x_1 y_1^T}{y_1^T x_1} \qquad (7.12.14)$$

<u>Voorbeeld</u>:

$$A = \begin{pmatrix} .2880 & .1360 & -.7120 \\ .2880 & .1360 & -.5120 \\ .5360 & .3920 & -.8640 \end{pmatrix}$$

Startend met $x^{(0)} = \begin{pmatrix} .5774 \\ .5774 \\ .5774 \end{pmatrix}$ vindt men $x^{(3)}$ en verder:

$x^{(3)}$	$Ax^{(3)}$	$x^{(4)}$	$Ax^{(4)}$	$x^{(5)}$	$Ax^{(5)}$
.7295	-.1156	-.7821	-.6743	-.5156	.3259
.4224	-.0080	-.0541	-.5502	-.4207	.1766
.5381	.0917	.6204	-.9764	-.7466	.2038
			1.3079		.4230

Toepassing van de vierkantsvergelijking op $x^{(3)}$, $Ax^{(3)}$, $|Ax^{(3)}|.Ax^{(4)}$ geeft:

$x^{(3)}$	$Ax^{(3)}$	$\|Ax^{(3)}\|.Ax^{(4)}$	
.7295	-.1156	-.0997	
.4224	-.0080	-.0813	$\lambda_1 = -.1999 +.4000i$
.5381	.0917	-.1443	$\lambda_2 = -.1999 -.4000i$

x_1	$x_1^* = \dfrac{x_1}{\sqrt{x_1^T x_1}}$	Ax_1^*	$\lambda_1 x_1^*$
-.1228-.0462i	-.6040-.2272i	.2119-.1961i	.2116-.1962i
-.0829-.0032i	-.4078-.0157i	.0879-.1600i	.0878-.1600i
-.1260+.0367i	-.6198+.1805i	.0519-.2839i	.0516-.2840i

Gaat men uit van $x^{(4)}$ enz., dan vindt men:

$\lambda_1 = -.2001+.4000i$

x_1	x_1^*	Ax_1^*	$\lambda_1 x_1^*$
.2913-.2697i	.4736-.4385i	.0807+.2772i	.0806+.2772i
.1209-.2201i	.1966-.3578i	.1038+.1502i	.1038+.1502i
.0712-.3906i	.1158-.6350i	.2309+.1733i	.2308+.1734i

Op dezelfde wijze y_1 itererend vindt men:

$\lambda_1 = -.2000+.3999i$

y_1^*
.4996-.0651i
.4305+.0932i
-.6378+.3819i

Voor Re $\dfrac{\lambda_1}{y_1^T x_1} \cdot x_1 y_1^T$ vindt men:

$$\begin{pmatrix} .1502 & .0502 & -.3502 \\ .1401 & .0801 & -.2601 \\ .2702 & .1902 & -.4303 \end{pmatrix}$$

en

$$A_2 = \begin{pmatrix} -.0124 & .0356 & -.0116 \\ .0078 & -.0242 & -.0082 \\ -.0044 & .0116 & -.0034 \end{pmatrix}$$

Bij A_2 vindt men $\lambda_3 = -.0400$

x_3	y_3
.8025	-.3025
-.5350	.9050
.2650	-.3000 .

7.13. Extrapolatie volgens Aitken

Heeft men een rij schattingen van een functiewaarde zodanig, dat men weet, dat de fout in iedere schatting een constant aantal malen de fout in de vorige is, dan kan men op eenvoudige wijze de fout elimineren. Laten gegeven zijn: $x_i = x + k^i d$, dus

$$x_0 = x + d$$
$$x_1 = x + kd$$
$$x_2 = x + k^2 d.$$

Maakt men differenties, dan vindt men:

$$\delta_{\frac{1}{2}} = (k-1)d; \quad \delta_{3/2} = k(k-1)d; \quad \delta_1^2 = (k-1)^2 d$$

en dus $$\frac{\delta_{\frac{1}{2}}\delta_{3/2}}{\delta_1^2} = kd \quad \text{en} \quad x = x_1 - \frac{\delta_{\frac{1}{2}}\delta_{3/2}}{\delta_1^2} \quad .$$

Met voorwaartse en teruglopende differenties krijgt men dergelijke formules:

$$\left.\begin{aligned} x &= x_0 - \frac{(\Delta_0)^2}{\Delta_0^2} \\ x &= x_1 - \frac{\delta_{\frac{1}{2}}\delta_{3/2}}{\delta_1^2} \\ x &= x_2 - \frac{(\nabla_2)^2}{\nabla_2^2} \end{aligned}\right\} \qquad (7.13.1)$$

en ook $$x = x_1 - \frac{\Delta_0 \nabla_0}{\delta_1^2}$$

Deze extrapolatie kunnen wij toepassen bij het itereren van eigenvectoren: wij weten immers, dat althans na enige tijd de fout bestaat (nagenoeg) alleen) uit de eerstvolgende eigenvector en dat deze met een constante (de verhouding van twee eigenwaarden) vermenigvuldigd wordt:

Wij zullen dit toepassen op het voorbeeld van pag. 44:

$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
.8900	.8759	.8734
-.2521	-.2242	-.2192
-.3800	-.4273	-.4349 .

Voor de geextrapoleerde elementen vindt men:

$$.8900 - \frac{4\ .25}{116} = .8729$$

$$-.2521 + \frac{279.50}{229} = -.2181$$

$$-.3800 - \frac{473.76}{397} = -.4364 \ ,$$

hetgeen met $x^{(7)}$ overeenstemt.
Men moet natuurlijk oppassen als δ_1^2 te klein wordt!

7.14. Alternatief om eigenwaarden te bepalen (squaring the matrix)

Bij deze methode bepaalt men A, A^2, A^4, A^8,... enz.
Nu volgt uit $A = X\Lambda Y^T = X\Lambda X^{-1}$ dat

$$A^2 = X\Lambda X^{-1} X\Lambda X^{-1} = X\,\Lambda^2 X^{-1}$$

A^2 heeft dus λ_i^2 en A^k heeft λ_i^k tot eigenwaarden. Als nu $|\lambda_1|>|\lambda_i|$, dan heeft, als k groot genoeg is, A^k één eigenwaarde λ_1^k en alle andere zijn ten opzichte hiervan 0.

$$A^k = \begin{pmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & & \vdots \\ x_1^{(n)} & \dots & x_n^{(n)} \end{pmatrix} \begin{pmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{pmatrix} \begin{pmatrix} y_1^{(1)} & \dots & y_1^{(n)} \\ \vdots & & \vdots \\ y_n^{(1)} & & y_n^{(n)} \end{pmatrix}$$

$$\approx \begin{pmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & & \vdots \\ x_1^{(n)} & \dots & x_n^{(n)} \end{pmatrix} \begin{pmatrix} \lambda_1^k y_1^{(1)} & \dots & \lambda_1^k y_1^{(n)} \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix} =$$

$$= \lambda_1^k \begin{pmatrix} x_1^{(1)}y_1^{(1)} & x_1^{(1)}y_1^{(2)} & \ldots & x_1^{(1)}y_1^{(n)} \\ x_1^{(2)}y_1^{(1)} & x_1^{(2)}y_1^{(2)} & \ldots & \vdots \\ \vdots & & & \vdots \\ x_1^{(n)}y_1^{(1)} & \ldots\ldots & & x_1^{(n)}y_1^{(n)} \end{pmatrix}$$

Heeft men A^{k_1} en A^{k_2}, dan vindt men een schatting voor $\lambda_1^{k_2-k_1}$ uit de verhouding van overeenkomstige elementen van A^{k_1} en A^{k_2}. Iedere rij van A^k nadert tot de eigenrij van A. Op gelijke wijze kan men de eigenkolom bepalen. Om hogere eigenwaarden te bepalen moet men teruggaan tot A en deze defleren: de matrix A^k (k groot) heeft geen andere eigenwaarde dan λ_1^k !

<u>Voorbeeld:</u>

$$A = \begin{pmatrix} 261 & -70 & -115 \\ -70 & 36 & 10 \\ -115 & 10 & 101 \end{pmatrix}$$

$$A^2 = \begin{pmatrix} 86246 & -21940 & -42370 \\ -21940 & 6296 & 9420 \\ -2330 & 9420 & 23526 \end{pmatrix}$$

$$10^{-5}A^4 = \begin{pmatrix} 97116 & -24291 & -48533 \\ -24291 & 6097 & 12096 \\ -48533 & 12096 & 24340 \end{pmatrix}$$

$$10^{-15}A^8 = \begin{pmatrix} 123770 & -30942 & -61884 \\ -30942 & 7735 & 15471 \\ -61884 & 15471 & 30942 \end{pmatrix}$$

Wij vinden een benadering voor λ_1 uit $\lambda_1^4 = 10^{10} \cdot \frac{123770}{97116}$.

$\lambda_1 = 336{,}0$

en $y_1 = x_1 = \begin{pmatrix} .8729 \\ -.2182 \\ .4364 \end{pmatrix}$, hetgeen met pag. 44 overeenstemt.

7.15. Convergentie van iteratie van lineaire stelsels

In 7.6 hebben wij de iteratiemethoden van Gauss en Seidel behandeld. Het ging om het oplossen van het stelsel:

$$Ax = b \tag{7.15.1}$$

waarbij $|A| \neq 0$.

De matrix A wordt bij genoemde processen gesplitst in twee matrices D + C en wel is bij de methode van Gauss D de diagonaal van A en C de rest en is bij de methode van Seidel D de beneden-driehoek + hoofddiagonaal van A.

$$Ax = (D+C)x = b \tag{7.15.2}$$

Als iteratie-formule wordt gebruikt:

$$Dx_k = b - Cx_{k-1}$$

of

$$x_k = D^{-1}b - D^{-1}Cx_{k-1} \tag{7.15.3}$$

Hierin zijn x_{k-1} en x_k resp. de k-1^e en k^e approximaties. Stel nu dat wij beginnen met $x_0 = 0$; wij vinden dan:

$$\begin{aligned}
x_0 &= 0 \\
x_1 &= D^{-1}b \\
x_2 &= (I-D^{-1}C)D^{-1}b \\
x_3 &= (I-D^{-1}C + (D^{-1}C)^2)D^{-1}b \\
&\;\;\vdots \\
x_k &= (I-D^{-1}C + \ldots + (-)^{k+1}(D^{-1}C)^{k-1})D^{-1}b \;.
\end{aligned}$$

Deze rij x_k convergeert slechts als alle eigenwaarden van $D^{-1}C$ absoluut kleiner dan 1 zijn.

In het algemeen is het moeilijk om voor een matrix A aan te geven of alle eigenwaarden van $D^{-1}C$, waarbij D diagonaal of beneden driehoek van A is, kleiner dan 1 zijn.

Voor het speciale geval dat voor iedere i geldt:

$$|a_{ii}| > \sum_{\substack{k=1 \\ k \neq i}}^{n} |a_{ik}| \tag{7.15.4}$$

zullen wij aantonen dat voor het geval dat D de diagonaal van A is alle eigenwaarden van $D^{-1}C < 1$ zijn.

Stel daartoe, dat wij een eigenvector x van $D^{-1}C$ hebben en laat het absoluut grootste element hiervan zijn x_k. Wij weten dan:

$$D^{-1} = \begin{pmatrix} \frac{1}{a_{11}} & & \\ & \ddots & \\ & & \frac{1}{a_{nn}} \end{pmatrix} \quad D^{-1}C = \begin{pmatrix} 0 & \frac{a_{12}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & 0 & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{a_{n1}}{a_{nn}} & \frac{a_{n2}}{a_{nn}} & \cdots & 0 \end{pmatrix}$$

en $D^{-1}Cx = \lambda x$,

wanneer λ de bij x behorende eigenwaarde is.

Wij bepalen nu het k^e element van x:

$$\left(\frac{a_{k1}}{a_{kk}} + \frac{a_{k2}}{a_{kk}} + \ldots + \frac{a_{k_1k-1}}{a_{kk}} + \ldots + \frac{a_{kn}}{a_{kk}}\right) \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ \vdots \\ x_n \end{pmatrix} = \sum_{\substack{i=1 \\ i\neq k}}^{n} \frac{a_{ki}}{a_{kk}} x_i =$$

$$\leqslant \sum_{\substack{i=1 \\ i\neq k}}^{n} \left|\frac{a_{ki}}{a_{kk}}\right| |x_i| \leqslant |x_k| \sum_{\substack{i=1 \\ i\neq k}}^{n} \left|\frac{a_{ki}}{a_{kk}}\right| .$$

En dit is wegens (7.15.4) kleiner dan $|x_k|$. Dus $|\lambda| < 1$.

8.1. Numeriek oplossen van gewone differentiaalvergelijkingen

Inleiding

Onder een gewone differentiaalvergelijking verstaan wij een betrekking tussen een onafhankelijke variabele x, een afhankelijke variabele y(x) en een of meer afgeleiden van y(x):

$$F(x,y,\frac{dy}{dx},\dots,\frac{d^n y}{dx^n}) = 0 \qquad (8.1.1)$$

Onder een oplossing van de differentiaalvergelijking (8.1.1) verstaan wij een y(x) zodat (8.1.1) geldt. Onder de orde van een differentiaalvergelijking verstaan wij de orde van de hoogste voorkomende afgeleide. Is de vergelijking lineair in y en zijn afgeleiden, dan heet de vergelijking lineair.
Wij zullen ons allereerst bezighouden met de algemene differentiaalvergelijkingen van de 1e orde:

$$F(x,y,\frac{dy}{dx}) = 0 \qquad (8.1.2)$$

In de analyse wordt aangetoond dat de oplossing de gedaante

$$g(x,y,c) = 0 \qquad (8.1.3)$$

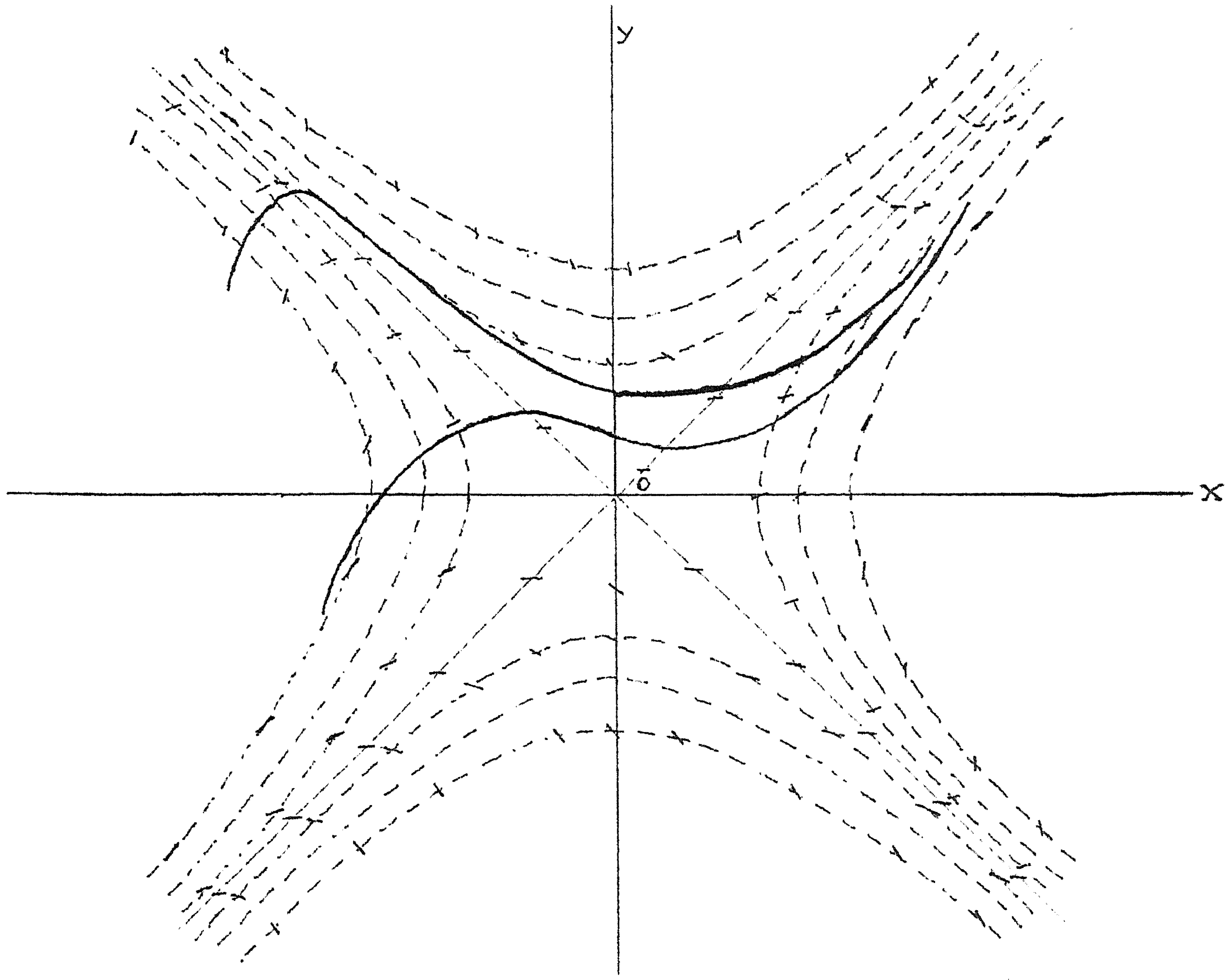
heeft: c is een willekeurige constante; iedere waarde hiervan geeft een andere oplossing.
In plaats van (8.1.2) zullen wij beschouwen:

$$\frac{dy}{dx} = f(x,y) \qquad (8.1.4)$$

Dit is geen beperking: als men $\frac{dy}{dx}$ niet analytisch kan oplossen, dan kan men immers bij gegeven x en y, $\frac{dy}{dx}$ uit (8.1.2) op een of andere wijze numeriek bepalen.

Een eenvoudige manier om inzicht te krijgen in de gedaante van de oplossingen voor een differentiaalvergelijking is het tekenen van de isoklinen (lijnen van gelijke hellingen) f(x,y) = const. Op een dergelijke lijn heeft $\frac{dy}{dx}$ steeds dezelfde waarde. Een manier van grafisch oplossen is: start in een punt van een isokline en ga met een helling $\frac{dy}{dx}$ verder totdat een nieuwe isokline bereikt wordt, ga verder met de nieuwe helling enz.

Beter doet men door tussen twee isoklienen een gemiddelde helling te nemen. Wij illustreren deze methode aan $y' = x^2 - y^2$. De isoklienen zijn $x^2 - y^2 = C$, d.w.z. gelijkzijdige hyperbolen. De lijnen $x = \pm y$ geven $\frac{dy}{dx} = 0$.



Men ziet dat met ieder startpunt een nieuwe oplossing correspondeert. Men kan een oplossing van (8.1.4) vastleggen door het geven van x_o en $y_o = y(x_o)$, de zg. beginvoorwaarde.

8.2. Taylor-ontwikkeling

$$\begin{aligned} y' &= f(x,y) \\ y(x_o) &= y_o \end{aligned} \tag{8.2.1}$$

Wij willen nu weten $y(x_o+h_o)$. Ontwikkel in Taylor-reeks:

$$y(x_o+h_o) = y(x_o) + y'(x_o)+h_o + \frac{y''(x_o)}{2} h_o^2 + \frac{y'''(x_o)}{6} h_o^3 + \ldots \tag{8.2.2}$$

Nu is:

$y(x_0) = y_0$

$y'(x_0) = f(x_0,y_0)$ volgens (8.2.1)

$y''(x_0) = f_x(x_0,y_0) + f_y(x_0,y_0)\,\frac{dy(x_0)}{dx} =$

$\qquad f_x(x_0,y_0) + f_y(x_0,y_0)f(x_0,y_0)$

$y\;\,(x_0) = f_{xx}(x_0,y_0) + 2f_{xy}(x_0,y_0)f(x_0,y_0) + f_{yy}(x_0,y_0)f^2(x_0,y_0) +$

$\qquad + f_y(x_0,y_0)f_x(x_0,y_0) + f_y^2(x_0,y_0)f(x_0,y_0)$.

Zo kan men verder gaan. In theorie kan men dus als x_0,y_0 gegeven zijn de waarden van de afgeleiden in het punt x_0 bepalen. Berekent men er een aantal van, b.v. n, dan kan men een h kiezen zo groot, dat de fout verwaarloosd kan worden die men maakt door te stellen:

$$y(x_0+h) = y(x_0) + \sum_{1}^{n} \frac{h_0^k}{k!}\, y^{(k)}(x_0) \qquad (8.2.3)$$

Vervolgens berekent men in $x_0 + h_0$ de afgeleiden en berekent $y(x_0+h_0+h_1)$, waarbij h_1 niet gelijk aan h_0 behoeft te zijn.

<u>Voorbeeld</u>

$y' = -2xy^2 \quad y(0) = 1 \quad$ oplossing $\quad y = \frac{1}{1+x^2}$.

$y'' = -2y^2-4xyy'$

x	y	y'	y''/2	fout
.0	1.000	.000	-1.000	0
.1	.990	-.196	- .941	0
.2	.961	-.369	- .782	-1
.3	.916	-.503	- .562	-1
.4	.860	-.592	- .332	-2
.5	.798	-.637	- .129	-2
.6	.733	-.645	+ .030	-2
.7	.669	-.627	+ .139	-2
.8	.608	-.592	+ .206	-2
.9	.551	-.546	+ .238	-1
1.0	.499			-1

Ter controle kan men ook een stap terug doen: h door -h vervangen:

$f(4) = -798.637 \times -.1 - .129 \times .01 = .860$.

Neemt men meer termen van de Taylor-reeks, dan kan men grotere intervallen nemen. De berekening van de differentiaalquotiënten wordt gecompliceerd: de methode is trouwens alleen te gebruiken als $f(x,y)$ een gemakkelijk te differentiëren functie is. Een voordeel is, dat het interval h niet gelijk hoeft te blijven. Men gebruikt de ontwikkeling dikwijls om een integratie van een differentiaalvergelijking te starten: men berekent dan b.v. $x_o \pm h$, $x_o \pm 2h$, $x_o \pm 3h$ uit de reeks in $x = x_o$.

De simpelste formule die men kan maken is:

$$y(x+h) = y(x) + hy'(x) \tag{8.2.4}$$

waarbij termen van orde h^2 verwaarloosd zijn.

Of als $y(x+nh) = y(x_n) = y_n$:

$$y_{n+1} = y_n + hy'_n + O(h^2) \tag{8.2.5}$$

Neem nu:

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2} y''_n + \frac{h^3}{6} y'''_n + O(h^4)$$

$$y_{n-1} = y_n - hy'_n + \frac{h^2}{2} y''_n - \frac{h^3}{6} y'''_n + O(h^4)$$

Aftrekken geeft:

$$y_{n+1} = y_{n-1} + 2hy'_n + O(h^3) \tag{8.2.6}$$

(8.2.6) geeft een kleinere fout dan (8.2.5).

Toepassen van (8.2.6) op het voorbeeld $y' = -2xy^2$ geeft hetzelfde resultaat als boven; toepassen van (8.2.5) geeft grotere fout:

(8.2.6)

x	y	y'	fout	y	y'	fout
.0	1.000	.000	0	1.000	.000	0
.1	.990	-.196	0	1.000	-.200	10
.2	.961	-.369	-1	.980	-.384	18
.3	.916	-.503	-1	.942	-.532	25
.4	.860	-.592	-2	.889	-.632	27
.5	.798	-.637	-2	.826	-.682	26
.6	.733	-.645	-2	.758	-.690	23
.7	.669	-.627	-2	.689	-.665	18
.8	.608	-.592	-2	.622	-.619	12
.9	.551	-.546	-1	.560	-.564	8
1.0	.499		-1	.504		4

Het boven geschetste grafische procédé komt neer op het toepassen van (8.2.5).
Het starten van (8.2.6) behoeft nadere toelichting: men heeft in (8.2.6) nodig twee punten y_n, x_{n-1}: dus naast x_o ook x_1 en y_1; y_1 moet men b.v. met de Taylor-reeks bepalen: dit moet goed gedaan worden, d.w.z. met grotere precisie.

8.3. Fout-opbouw

In deze paragraaf zullen wij de fout-opbouw behandelen die optreedt bij het toepassen van (8.2.5) en (8.2.6). Door afrondingen en afbreken van Taylor-reeks maakt men fouten: of deze zich voortplanten of uitsterven zal onderzocht worden.
Allereerst:

$$y_{n+1} = y_n + hy'_n \tag{8.3.1}$$

of met $y' = f(x,y)$: $\quad y_{n+1} = y_n + hf(x_n,y_n)$ (8.3.2)

Laat z_n de echte oplossing zijn, dan geldt:

$$z_{n+1} = z_n + hf(x_n,z_n) + T_n \tag{8.3.3}$$

waarbij T_n de afbreekfout is. Stel $z_n = y_n + \varepsilon_n$, dan geeft aftrekken van (8.3.2) van (8.3.3):

$$\varepsilon_{n+1} = \varepsilon_n + h\{f(x_n,z_n) - f(x_n,y_n)\} + T_n \tag{8.3.4}$$

Volgens de middelwaarde-stelling is:

$$f(x_n,z_n) - f(x_n,y_n) = (z_n-y_n)f_y(x_n,\eta_n) = \varepsilon_n g_n$$
of $$\varepsilon_{n+1} = \varepsilon_n + h\varepsilon_n g_n + T_n \qquad (8.3.5)$$

Om iets te kunnen doen nemen wij aan, dat zowel g_n als T_n onafhankelijk zijn van n.
Wij hebben dan:

$$\varepsilon_{n+1} = \varepsilon_n(1+hg) + T \qquad (8.3.6)$$

Om deze recursiebetrekking op te lossen nemen wij eerst de homogene:

$$\varepsilon_{n+1} = \varepsilon_n(1+hg)$$

en vinden

$$\varepsilon_n = (1+hg)^n \ .$$

Substitutie in (8.3.6):van $\varepsilon_n = (1+hg)^n + C$ geeft:

$$(1+hg)^{n+1} + C = (1+hg)\{(1+hg)^n + C\} + T$$

of

$$C = -\frac{T}{hg} \ .$$

Dus $$\varepsilon_n = (1+hg)^n - \frac{T}{hg} \ , \qquad (8.3.7)$$

waarbij g een of andere waarde van f_y is. Stel $h > 0$. Tenzij $g < 0$ is, is $(1+hg) > 1$ en groeit ε_n exponentieel: dit is een eigenschap van de differentiaalvergelijking waaraan niets te veranderen is. Is $g < 0$, dan kan men h te groot nemen, nl. $h > |\frac{1}{g}|$.
Voor $0 < h < |\frac{1}{g}|$ groeien de fouten niet aan.

Nu $$y_{n+1} = y_{n-1} + 2hy'_n$$

$$z_{n+1} = z_{n-1} + 2hz'_n + T_n$$

en

$$\varepsilon_{n+1} = \varepsilon_{n-1} + 2h\{f(x_n,z_n) - f(x_n,y_n)\} + T_n$$

of $\quad \varepsilon_{n+1} = \varepsilon_{n-1} + 2h\varepsilon_n g_n + T_n$ met $g_n = f_y(x_n, y_n)$

of $\quad \varepsilon_{n+1} - 2hg\varepsilon_n - \varepsilon_{n-1} = T. \qquad (8.3.8)$

Dit is een tweede orde recursie-vergelijking.

Homogene vergelijking $\varepsilon_{n+1} - 2hg\varepsilon_n - \varepsilon_{n-1} = 0$.

Oplossing proberen $\varepsilon_n = r^n$: r moet voldoen aan:

$$r^2 - 2hgr - 1 = 0 \ .$$

Laten dit zijn r_1 en r_2. Het product is -1 dus tenzij $|r_1| = |r_2| = 1$ is een wortel absoluut >1.

Wij vinden $\varepsilon_n = \alpha r_1^n + \beta r_2^n$ want beide functies r_1^n en r_2^n voldoen aan de homogene vergelijking, dus ook een lineaire combinatie.

Nu de oplossing van (8.3.8):

$$\varepsilon_n = \alpha r_1^n + \beta r_2^n + C$$

$$\alpha r_1^{n+1} + \beta r_2^{n+1} + C - 2hg(\alpha r_1^n + \beta r_2^n + C) - \alpha_1 r_1^{n-1} - \beta r_2^{n-1} - C = T$$

of $C = -\dfrac{T}{2hg}$.

$$\therefore \qquad \varepsilon_n = \alpha r_1^n + \beta r_2^n - \frac{T}{2hg} \ . \qquad (8.3.9)$$

Nu is het triest: wij hebben immers boven gevonden, dat of r_1 of r_2 absoluut groter dan 1 is, onafhankelijk van h of g: ε_n groeit dus altijd exponentieel met toenemende n. (8.2.6) is essentieel een instabiele recursie-vergelijking. Nu behoeft men voor kleine waarden van n geen last te hebben, en dan heeft men betere precisie (of groter interval) dan bij (8.2.5).

8.4. Methode van Runge-Kutta

Deze methode waaraan behalve de namen van Runge, Kutta, Heun en Gill verbonden zijn stelt ons in staat o.a. vergelijkingen van het type

$$y' = f(x,y) \qquad (8.4.1)$$

op te lossen.

Als gegeven is $y(x) = y$, dan zullen wij trachten $y(x+h)$ te schrijven als lineaire combinatie van y en van waarden $f(x_i,y_i)$ zodanig dat (x_i,y_i) punten zijn in de buurt van (x,y), b.v.:

$$y(x+h) = y + \alpha_o k_o + \alpha_1 k_1 + \dots + \alpha_p k_p \qquad (8.4.2)$$

waarin

$$\begin{aligned} k_o &= hf(x,y) \\ k_1 &= hf(x+\mu_1 h, y+\lambda_{10}k_o) \\ k_2 &= hf(x+\mu_2 h, y+\lambda_{20}k_o+\lambda_{21}k_1) \\ &\vdots \\ k_p &= hf(x+\mu_p h, y+\sum_{i=0}^{p-1}\lambda_{pi}k_i) \end{aligned}$$

Wij zullen $p = 2$ nemen:

$$y(x+h) = y + \alpha_o k_o + \alpha_1 k_1 + \alpha_2 k_2 \qquad (8.4.3)$$

$$\begin{aligned} k_o &= hf(x,y) \\ k_1 &= hf(x+\mu_1 h, y+\lambda_1 k_o) \\ k_2 &= hf(x+\mu_2 h, y+\lambda_2 k_o+\lambda_3 k_1)\ . \end{aligned}$$

Wij hebben nu als parameters $\alpha_o, \alpha_1, \alpha_2, \mu_1, \mu_2, \lambda_1, \lambda_2, \lambda_3$, dus 8 in totaal en zullen trachten deze zo te bepalen dat $y(x+h)$ gegeven door (8.4.3) als men deze naar h ontwikkelt in zoveel mogelijk termen overeenstemt met de Taylor-ontwikkeling:

$$y(x+h) = y + hy' + \tfrac{1}{2}h^2y'' + \tfrac{1}{6}h^3y''' + \dots \qquad (8.4.4)$$

Wij weten

$$\begin{aligned} y' &= f \qquad\qquad f \equiv f(x,y) \\ y'' &= f_x + f_y f \\ y''' &= f_{xx} + 2f_{xy}f + f_{yy}f^2 + f_x f_y + f_y^2 f \end{aligned}$$

Wij moeten nu (8.4.3) naar h ontwikkelen:

$$\begin{aligned} k_o &= hf \\ k_1 &= hf(x+\mu_1 h, y+\lambda_1 fh) = fh + (\mu_1 f_x + \lambda_1 f_y f)h^2 + \\ &\quad + (\tfrac{1}{2}\mu_1^2 f_{xx} + \mu_1\lambda_1 f_{xy}f + \tfrac{1}{2}\lambda_1^2 f_{yy}f^2)h^3 + \dots \qquad (8.4.5) \end{aligned}$$

Hierbij is gebruik gemaakt van de Taylor-ontwikkeling van een functie van twee variabelen:

$$f(x+h\ ,y+k) = f(x,y) + \sum_1^{\infty}\frac{1}{n!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^n f\ ,$$

waarbij verstaan wordt onder:

$$n = 1 \qquad \left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)f = hf_x + kf_y$$

$$n = 2 \qquad \left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^2 f = \left(h^2\frac{\partial^2}{\partial x^2}+2hk\frac{\partial^2}{\partial x\,\partial y}+k^2\frac{\partial^2}{\partial y^2}\right)f =$$

$$= h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy}$$

enz.

Vervolgens moet k_2 ontwikkeld worden:

$$k_2 = hf$$

$$k_2 = hf(x+\mu_2 h, y+\lambda_2 k_0+\lambda_3 k_1) =$$

$$= h\left[f+\mu_2 f_x h+(\lambda_2 k_0+\lambda_3 k_1)f_y+\tfrac{1}{2}\mu_2^2 f_{xx}h^2+\mu_2 h(\lambda_2 k_0+\lambda_3 k_1)f_{xy} +\right.$$

$$\left.+ \tfrac{1}{2}(\lambda_2 k_0+\lambda_3 k_1)^2 f_{yy}+\ldots\right] =$$

$$= hf + \mu_2 f_x h^2 + h\left\{\lambda_2 hf+\lambda_3 \left(hf+(\mu_1 f_x+\lambda_1 f_y f)h^2\right)\right\}f_y + \tfrac{1}{2}\mu_2^2 f_{xx}h^3 +$$

$$+ \mu_2 h^2(\lambda_2 hf+\lambda_3 hf)f_{xy}+\tfrac{1}{2}h(\lambda_2 hf+\lambda_3 hf)^2 f_{yy} + \ldots$$

$$\therefore\ k_2 = fh + \left\{\mu_2 f_x+(\lambda_2+\lambda_3)ff_y\right\}h^2 + \left\{\lambda_3(\mu_1 f_x+\lambda_1 ff_y)f_y +\right.$$

$$\left.+ \mu_2(\lambda_2+\lambda_3)ff_{xy} + \tfrac{1}{2}(\lambda_2+\lambda_3)^2 f_{yy}f^2 + \tfrac{1}{2}\mu_2^2 f_{xx}\right\}h^3 + \ldots$$

Voor $y(x+h) = y + \alpha_0 k_0 + \alpha_1 k_1 + \alpha_2 k_2$

vinden wij dus als coëfficiënt van:

h^0: y

h^1: $\alpha_0 f + \alpha_1 f + \alpha_2 f$

h^2: $\alpha_1(\mu_1 f_x+\lambda_1 f_y f) + \alpha_2(\mu_2 f_x+(\lambda_2+\lambda_3)f_y f)$

h^3: $\alpha_1(\tfrac{1}{2}\mu_1^2 f_{xx}+\mu_1\lambda_1 f_{xy}f+\tfrac{1}{2}\lambda_1^2 f_{yy}f^2) + \alpha_2\{\lambda_3(\mu_1 f_x+\lambda_1 ff_y)f_y +$

$\quad + \mu_2(\lambda_2+\lambda_3)f_{xy}f + \tfrac{1}{2}(\lambda_2+\lambda_3^2)f_{yy}f^2 + \tfrac{1}{2}\mu_2^2 f_{xx}\}$.

Vergelijking met (8.4.4) geeft de vergelijkingen:

1) $\alpha_0 + \alpha_1 + \alpha_2 = 1$
2) $\alpha_1\mu_1 + \alpha_2\mu_2 = \frac{1}{2}$
3) $\alpha_1\lambda_1 + \alpha_2(\lambda_2+\lambda_3) = \frac{1}{2}$
4) $\frac{1}{2}\alpha_1\mu_1^2 + \frac{1}{2}\alpha_2\mu_2^2 = \frac{1}{6}$
5) $\alpha_1\mu_1\lambda_1 + \alpha_2\mu_2(\lambda_2+\lambda_3) = \frac{1}{3}$
6) $\frac{1}{2}\alpha_1\lambda_1^2 + \frac{1}{2}\alpha_2(\lambda_2+\lambda_3)^2 = \frac{1}{6}$
7) $\alpha_2\lambda_3\mu_1 = \frac{1}{6}$
8) $\alpha_2\lambda_1\mu_3 = \frac{1}{6}$

Hier staan 8 vergelijkingen, schijnbaar nodig en voldoende om de 8 onbekenden op te lossen. Echter moet opdat (2) en (3) een andere dan nuloplossing hebben $\mu_1(\lambda_2+\lambda_3) = \mu_2\lambda_1$ zijn.
Dus in plaats van (3) komt:

3a) $\mu_1(\lambda_2+\lambda_3) = \mu_2\lambda_1$

Maar (3a) impliceert dat (4) en (5) en (6) afhankelijk zijn: (5) en (6) vervallen dus; wij hebben dus slechts 6 vergelijkingen over. Kiezen wij μ_1 en μ_2 vrij, dan vindt men:

$$\alpha_1 = \frac{\frac{\mu_2}{2} - \frac{1}{3}}{\mu_1(\mu_2-\mu_1)} \; ; \quad \alpha_2 = \frac{\frac{\mu_1}{2} - \frac{1}{3}}{\mu_1(\mu_1-\mu_2)} \; ; \quad \lambda_3 = \frac{1}{6\mu_1\alpha_2} \; ;$$

en $\lambda_1 = \mu_1$; $\lambda_2 = \mu_2 - \lambda_3$; $\alpha_0 = 1 - \alpha_1 - \alpha_2$.

Nu kan men aantonen dat het niet mogelijk is μ_1 en μ_2 zo te kiezen dat ook de term met h^4 overeenstemt met die van de Taylor-reeks. De fout is dus steeds $O(h^4)$, wat μ_1 en μ_2 zijn.
Laten wij eenvoudige waarden kiezen:

$\mu_1 = \frac{1}{2}$, $\mu_2 = 1$: $\alpha_1 = \frac{2}{3}$; $\alpha_2 = \frac{1}{6}$; $\alpha_0 = \frac{1}{6}$; $\lambda_1 = \frac{1}{2}$; $\lambda_2 = -1$; $\lambda_3 = 2$

en de formules worden:

$$y(x+h) = y + \frac{1}{6}(k_0+4k_1+k_2) + O(h^4) \qquad (8.4.5)$$

$k_0 = hf(x,y)$
$k_1 = hf(x+\frac{1}{2}h, y+\frac{1}{2}k_0)$
$k_2 = hf(x+h, y-k_0+2k_1)$

of $\mu_1 = \frac{1}{3}, \mu_2 = \frac{2}{3}$: $\alpha_0 = \frac{1}{4}$; $\alpha_1 = 0$; $\alpha_2 = \frac{3}{4}$; $\lambda_1 = \frac{1}{3}$; $\lambda_2 = 0$;
$\lambda_3 = \frac{2}{3}$

$$y(x+h) = y + \frac{1}{4}(k_0+3k_2) + O(h^4) \qquad (8.4.6)$$

$$k_0 = hf(x,y)$$
$$k_1 = hf(x+\tfrac{1}{3}h, y+\tfrac{1}{3}k_0)$$
$$k_2 = hf(x+\tfrac{2}{3}h, y+\tfrac{2}{3}k_1)$$

Wenst men een fout $O(h^3)$ toe te laten, dan stelle men $\alpha_2 = 0$:

$$\alpha_0 + \alpha_1 = 1$$
$$\alpha_1\mu_1 = \tfrac{1}{2}$$
$$\alpha_1\lambda_1 = \tfrac{1}{2} \; .$$

Stel $\alpha_1 = c$; $\alpha_0 = 1 - c$; $\mu_1 = \frac{1}{2c}$; $\lambda_1 = \frac{1}{2c}$.
Nu is c niet zo te kiezen, dat de h^3-term klopt. Kies dan $c = \frac{1}{2}$:

$$\alpha_1 = \tfrac{1}{2} = \alpha_0 \qquad \mu_1 = \lambda_1 = 1 \; .$$

$$y(x+h) = y + \tfrac{1}{2}(k_0+k_1) + O(h^3) \qquad (8.4.7)$$

$$k_0 = hf(x,y)$$
$$k_1 = hf(x+h, y+k_0) \; .$$

Als f onafhankelijk is van y, dan reduceert (8.4.5) tot de regel van Simpson en (8.4.7) tot de trapezium-regel.

Een voordeel van deze soort formules is dat men direct starten kan: men hoeft geen Taylor-reeks te bepalen, geen hogere afgeleiden uit te rekenen, enz. Verder kan men gemakkelijk (maar dit geldt ook voor de andere tot nu toe behandelde methoden) het interval te wijzigen. Is het berekenen van f veel werk, dan zal men een andere integratiemethode die later behandeld worden prefereren.

8.5. Hogere orde vergelijkingen

In het voorgaande zijn enige methoden behandeld voor het oplossen van differentiaalvergelijkingen van de 1e orde.
Heeft men een hogere orde vergelijking, dan kan men deze altijd schrijven als een stelsel simultane vergelijkingen van de 1e orde:

$$y^{(n)} = \mathbf{f}(y^{(n-1)},\ y^{(n-2)}, \ldots, y^{(1)},\ y,\ x) \qquad (8.5.1)$$

gaat door substitutie van

$$\begin{aligned} y &= y_o \\ y^{(1)} &= y_1 \\ y^{(2)} &= y_2 \\ y^{(n)} &= y_n \end{aligned}$$

over in het stelsel

$$\begin{aligned} y_o' &= y_1 \\ y_1' &= y_2 \\ &\vdots \\ y_{n-1}' &= y_n \\ y_n' &= f(y_{n-1}, y_{n-2}, \ldots, y_1, y_o, x) \end{aligned} \qquad (8.5.2)$$

Dit stelsel met de onbekenden $y_o \ldots y_n$ dient in overeenstemming met de gegeven randvoorwaarden te worden opgelost.
Men kan dat wederom doen met behulp van de Taylorreeks: men moet dan alle functies y_i ontwikkelen en steeds alle vergelijkingen een stap integreren; dan weer afgeleiden berekenen, enz.

Ook kan men Runge-Kutta - formules maken voor simultane vergelijkingen.

Deze methoden zijn zonder meer toepasbaar op stelsels vergelijkingen van minder eenvoudige aard dan (8.5.2). Daar zijn immers alle rechter leden op één na functies van een variabele. Het stelsel

$$
\begin{aligned}
y_o' &= f_o(y_o,\ldots,y_n,x)\\
y_1' &= f_1(y_o,\ldots,y_n,x)\\
&\vdots\\
y_n' &= f_n(y_o,\ldots,y_n,x)
\end{aligned}
\qquad (8.5.3)
$$

kan men op dezelfde wijze behandelen.
Tenslotte geven wij een formule van Runge-Kutta (vergelijk (8.4.5)) voor n+1 vergelijkingen:

$$
\begin{aligned}
y_i(x+h) &= y_i(x) + \tfrac{1}{6}(k_{oi}+4k_{1i}+k_{2i}) + O(h^4)\\
k_{oi} &= hf_i(y_o(x)\ldots y_n(x),x)\\
k_{1i} &= hf_i(y_o+\tfrac{1}{2}k_{oo},\ldots,y_n+\tfrac{1}{2}k_{on},x+\tfrac{1}{2}h)\\
k_{2i} &= hf_i(y_o+2k_{1o}-k_{oo},\ldots,y_n+2k_{1n}-k_{on},x+h)
\end{aligned}
\qquad (8.5.4)
$$

8.6. Tweede orde vergelijkingen

In deze en de volgende paragraaf zullen wij differentie-methoden behandelen. Allereerst de vergelijking van de tweede orde

$$y'' = f(x,y) \qquad (8.6.1)$$

Dit is een eenvoudiger type dan de algemenere $y'' = f(x,y,y')$.
De algemene lineaire tweede orde vergelijking

$$y'' + f(x)y' + g(x)y = h(x) \qquad (8.6.2)$$

kan men echter tot (8.6.1) herleiden, zodat (8.6.1) toch niet zo'n erg speciale vergelijking is. Stel $y = vz$, v en z functies van x, dan is

$y' = v'z + vz'$
$y'' = v''z + 2v'z' + vz''$

Dit gesubstitueerd in (8.6.2) geeft:

$$z''v + z'(2v'+fv) + z(v''+fv'+gv) = h \qquad (8.6.3)$$

Als wij nu $2v' = -fv$ kiezen, dan wordt (8.6.3) van de vorm (8.6.1); dus $\ln v = -2\int f\,dx + C$ of
$$v = C\,e^{-2\int f\,dx}.$$

Als $2v' = -fv$, dan $v'' = -\frac{1}{2}(f'v+fv') = \frac{1}{4}f^2v - \frac{1}{2}f'v$ en
$v'' + fv' + gv = -\frac{1}{4}f^2v - \frac{1}{2}f'v + gv.$

Dus (8.6.3) gaat over in:

$$z'' + z(g-\tfrac{1}{4}f^2-\tfrac{1}{2}f') = \frac{h}{v} \qquad (8.6.4)$$

van de vorm (8.6.1); z' komt niet meer voor.

Voorbeeld

Vergelijking van Bessel $y'' + \frac{1}{x}y' + (1-\frac{p^2}{x^2})y = 0$

$f = \frac{1}{x}$, $g = 1 - \frac{p^2}{x^2}$, $h = 0$

$v = 2^{-2\int \frac{1}{x}dx} = \frac{1}{\sqrt{x}}$.

Dus $y = \sqrt{x}\,z$ en voor z komt de vergelijking:

$$z'' + z(1-\frac{p^2}{x^2}+\frac{1}{4x^2}) = 0$$

Wij beperken ons dus tot het oplossen van

$$\begin{aligned} &y'' = f(x,y) \\ &y(x_o) = y_o \\ &y'(x_o) = y'_o \end{aligned} \qquad (8.6.5)$$

Stel $h^2y'' = Q$, waarin h het gekozen interval voorstelt, dan dus

$Q = h^2f(x,y)$

Stel nu dat wij de oplossing y gevonden hebben tot y_n. Wij kunnen dan de volgende differentietabel opstellen:

$$
\begin{array}{cccccccc}
x_{n-3} & y_{n-3} & \delta^{-2}_{n-3} & & Q_{n-3} & & \delta^{2}_{n-3} & \\
 & & & \delta^{-1}_{n-\frac{5}{2}} & & \delta_{n-\frac{5}{2}} & & \delta^{3}_{n-\frac{5}{2}} \\
x_{n-2} & y_{n-2} & \delta^{-2}_{n-2} & & Q_{n-2} & & \delta^{2}_{n-2} & \\
 & & & \delta^{-1}_{n-\frac{3}{2}} & & \delta_{n-\frac{3}{2}} & & \delta^{3}_{n-\frac{3}{2}} \\
x_{n-1} & y_{n-1} & \delta^{-2}_{n-1} & & Q_{n-1} & & \delta^{2}_{n-1} & \\
 & & & \delta^{-1}_{n-\frac{1}{2}} & & \delta_{n-\frac{1}{2}} & & \\
x_{n} & y_{n} & \delta^{-2}_{n} & & Q_{n} & & & \\
 & & & \delta^{-1}_{n+\frac{1}{2}} & & & & \\
x_{n+1} & & \delta^{-2}_{n+1} & & & & &
\end{array}
$$

Wij zouden nu y_{n+1} kunnen bepalen door met een teruglopende dubbele-integratie-formule een stap te integreren. Deze formule hebben wij in 4.5 niet gegeven, daar de convergentie der coëfficiënten slechts is:

$$y_{n+1} = \nabla^{-2}_n + \frac{1}{12}Q_n + \frac{1}{12}\nabla_n + \frac{19}{240}\nabla^2_n + \ldots \qquad (8.6.6)$$

Sneller convergeert de contrale formule (4.5.1):

$$y_{n+1} = \delta^{-2}_{n+1} + \frac{1}{12}Q_{n+1} - \frac{1}{240}\delta^2_{n+1} + \frac{31}{60480}\delta^4_{n+1} \qquad (8.6.7)$$

Het gebruiken van (8.6.7) wordt echter enigszins bemoeilijkt door het feit, dat behalve de kopterm geen der andere termen bekend is.

Een manier om deze moeilijkheid te overwinnen is extrapolatie van alles wat wij nodig hebben en onbekend is: dus Q_{n+1} en eventueel $^{2}_{n+1}$ enz. Dit kan zonder moeite als men een "gladde" differentietabel heeft. Met deze geschatte grootheden berekent men een voorlopige y_{n+1}; hieruit Q_{n+1} en de daaruitvolgende teruglopende differenties: is de Q_{n+1} veel verschillend van de schatting, dan berekent men weer y_{n+1}, enz. Men kieze h zo klein, dat men y_{n+1} direct goed vindt.
Wij hebben aangenomen, dat de integratie een eind geschied was.
Ons ontbreekt nog de start.
Twee methoden zullen wij behandelen:
1. Reeksontwikkeling. Men bepaalt met de reeksontwikkeling:

$$y(x) = y_o + y_o'(x-x_o) + \tfrac{1}{2}y_o''(x-x_o)^2 + \tfrac{1}{6}y_o'''(x-x_o)^3 + \ldots \qquad (8.6.8)$$

een aantal functiewaarden, b.v. $y(x_o \pm 3h)$, $y(x_o \pm 2h)$, $y(x_o \pm h)$ en eerste afgeleiden, en hieruit de bijbehorende waarden van Q en differenties.
Uit (8.6.7) kan men dan δ^{-2}_{n+1} bepalen en uit (4.4.5):

$$hy'_{n+\frac{1}{2}} = \delta^{-1}_{n+\frac{1}{2}} + \frac{1}{24}\,\delta_{n+\frac{1}{2}} - \frac{17}{5700}\,\delta^{3}_{n+\frac{1}{2}} \quad \ldots \qquad (8.6.9)$$

bepaalt men $\delta^{-1}_{n+\frac{1}{2}}$.
Als het goed is, moeten de eerste somfuncties differenties van de tweede zijn en de functies differenties van de eerste somfuncties zijn. Dit hoeft wegens afrondingen niet precies te kloppen. Het is echter wel nodig, bij een start, zeer precies te werk te gaan: een fout in de eerste somfunctie b.v. wordt lineair opgebouwd!

2. Iteratie. Men neemt als approximatie b.v. $y(x) = y_o + y_o'(x-x_o)$ en bepaalt y en Q weer voor een aantal basispunten. Met de differenties en (8.6.7) bepaalt men met behulp van de gegeven y_o δ_o^{-2} en met (4.4.4).

$$hy_o' = \mu\delta_o^{-1} - \frac{1}{12}\mu\delta_o + \frac{11}{720}\mu\delta_o^3 + \ldots \qquad (8.6.10)$$

kan men uit de gegeven y_o' de constante in de eerste somfunctie bepalen.

Daarna kan men met (8.6.7) integreren: men vindt dan een rij nieuwe waarden voor y. Men bepaalt weer Q en de constante in de somfuncties enz. totdat de zaak stabiel blijkt.

<u>Voorbeeld</u>

$y'' + xy = 0, \quad y(0) = 0; \quad y'(0) = 1$

$y = \sum a_k x^k; \qquad y'' = \sum_2 k(k-1)a_k x^{k-2} = -\sum_0 a_k x^{k+1}$

Dus $a_2 = 0$ en $a_{k+3} = \frac{-1}{(k+2)(k+3)}\, a_k$.

De randvoorwaarden geven:

$a_0 = 0$
$a_1 = 1$.

Wij vinden dus:

$$y(x) = x - \frac{1}{3.4}\, x^4 + \frac{1}{3.4.6.7}\, x^7 - \frac{1}{3.4.6.7.9.10}\, x^{10} + \ldots$$

$$y'(x) = 1 - \frac{1}{3}x^3 + \frac{1}{3.4.6}\, x^6 - \frac{1}{3.4.6.7.9}\, x^9 + \ldots$$

Hiermede berekenen wij:

x	y(x)	y'(x)
-.3	-.300675	
-.25		+1.005212
-.2	-.200133	
-.15		+1.001125
-.1	-.100008	
-.05		+1.000042
-.0	0	
+.05		+ .999958
+.10	+.099992	
+.15		+ .998875
+.20	+.199867	
+.25		+ .994795
+.30	+.299325	

Vervolgens berekenen wij $Q = -h^2 xy$ voor $x = -.3(.1)+.3$ en bepalen de differenties van Q.

Daarna bepalen wij:

δ_y^{-2} voor x = -.3(.1)+.3 met (8.6.7) en

δ_y^{-1} voor x = -.25(.1)+.25 met (8.6.9) .

Wij vinden dan de start:

x	y	δ^{-2}	δ^{-1}	Q		δ^2	δ^3
-.3	-.300675	-300601		-902			
			100500		502		
-.2	-.200133	-200100^5		-400		-202	
			100100		300		2
-.1	-.100008	-100000^5		-100		-200	
			100000		100		0
0	0	- 1		0		-200	
			100000		-100		0
.1	+.099992	+ 99999^5		-100		-200	
			99900		-300		2
.2	+.199867	+199899^5		-400		-198	
			99500		-498		
.3	+.299235	+299399		-898			
			98602				
.4		+398001					

Nu kan de integratie beginnen. Wij vinden:

x	y	δ^{-2}	δ^{-1}	Q		δ^2	δ^3	δ^4
0	.000000	- 1		0		-200		
			100000		- 100		0	
.1	.099992	99999^5		- 100		-200		[illegible]
			99900		- 300		2	
.2	.199867	199899^5		- 400		-198		1
			99500		- 498		3	
.3	.299325	299399		- 898		-195		2
			98602		- 693		5	
.4	.397869	398001		-1591		-190		6
			97011		- 883		11	
.5	.494807	495012		-2474		-179		5
			94537		-1062		16	
.6	.589255	589549		-3536		-163		4
			91001		-1225		19	
.7	.680154	680550		-4761		-144		8
			86240		-1369		27	
.8	.766280	766790		-6130		-117		6
			80110		-1486		33	
.9	.846266	846900		-7616		- 84		(9)
			72494		-1570		(42)	
1.0	.918629	919394		-9186		(-42)		

De getallen tussen haakjes zijn geschatte waarden.

De exacte oplossing van de vergelijking is

$$y(z) = \tfrac{1}{3}!\,3^{1/3}\;.\;\sqrt{x}\;J_{1/3}\left(\tfrac{2}{3}\,x^{2/3}\right),$$

waarbij $J_{1/3}(z)$ een getabelleerde Besselfunctie is. Men vindt dat onze functie y(x) overeenstemt in alle decimalen met de exacte waarden, behalve x = .4, waarvoor men vindt .397870.

Als alternatief itereren wij de start ook. Kies als beginschatting y = x; bepaal Q en differenties en bepaal de somfuncties. Wij vinden $10^6\,\delta_0^{-2} = -1$ en $10^6\mu\delta_0^{-1} = 100000$. De tabel ziet er dan als volgt uit:

x	y	δ^{-2}	δ^{-1}	Q	δ	δ^2
-.3	-.300000	-300601		-900		
			100500		500	
-.2	-.200000	-200101		-400		-200
			100100		300	
-.1	-.100000	-100001		-100		-200
			100000		100	
0	0	- 1		0		-200
			100000		-100	
+.1	+.100000	99999		-100		-200
			99900		-300	
+.2	+.200000	199899		-400		-200
			99500		-500	
+.3	+.300000	299399		-900		

Vervolgens integreren wij nieuwe y met (8.6.7) en vinden

x	y	δ^{-2}	δ^{-1}	Q	δ	δ^2
-.3	-.300675	-300601		-902		
			100500		502	
-.2	-.200134	-200101		-400		-202
			100100		300	
-.1	-.100008	-100001		-100		-200
			100000		100	
0	0	- 1		0		-200
			100000		-100	
.1	.099992	99999		-100		-200
			99900		-300	
.2	.199866	199899		-400		-198
			99500		-498	
.3	.299325	299399		-898		

hetgeen overeenstemt met de andere resultaten.

Tot nu toe hebben wij als integratieformule gebruikt (8.6.7):

$$y_n = \delta_n^{-2} + \frac{1}{12} Q_n - \frac{1}{240} \delta_n^2 + \frac{31}{60480} \delta_n^4 - \dots$$

Als wij hiervan de tweede differentie nemen vinden wij:

$$\delta^2 y_n = Q_n + \frac{1}{12} \delta^2 Q_n - \frac{1}{240} \delta^4 Q_n + \frac{31}{60480} \delta^6 Q_n - \dots \qquad (8.6.11)$$

Met (8.6.11) kunnen wij uit Q_n en geschatte differenties bepalen $\delta^2 y_n$ en $y_{n+1} = y_n + \delta_{n-\frac{1}{2}} + \delta_n^2$.

Bij gebruik van (8.6.11) in plaats van (8.6.7) behoeven wij dus niet Q maar slechts de differenties te extrapoleren. Echter is de zaak niet zo gunstig als het lijkt: uit Q en zijn differenties wordt $\delta^2 y_n$ bepaald met afrondingsfouten; deze worden in de y lineair opgebouwd, dit in tegenstelling tot (8.6.7) waar de somfuncties zonder afrondingsfouten uit Q werden opgebouwd.

8.7. Eerste orde vergelijkingen

Laat te bepalen zijn y(x) zodat

$$y' = f(x) \qquad (8.7.1)$$
$$y(x_o) = y_o$$

Stel $hy' = P$, dan is $P = hf(x)$ (8.7.2)

Laat bekend zijn y(x) tot x_n: y_n.
Wij zouden dan met (4.3.7) kunnen bepalen

$$y_{n+1} - y_n = \int_{x_n}^{x_{n+1}} f(x)dx = \frac{1}{h}\int_{x_n}^{x_{n+1}} P\, dx = (\nabla_{n+1}^{-1} - \nabla_n^{-1}) - \tfrac{1}{2}(f_{n+1} - f_n)$$

$$\frac{1}{12}(\nabla_{n+1} - \nabla_n) - \frac{1}{24}(\nabla_{n+1}^2 - \nabla_n^2) - \frac{19}{720}(\nabla_{n+1}^3 - \nabla_n^3) - \frac{3}{160}(\nabla_{n+1}^4 - \nabla_n^4) - \dots \qquad (8.7.3)$$

Wij kunnen nu de differenties met index n+1 uitdrukken in die met index n:

$$\nabla_{n+1}^k = \nabla_n^k + \nabla_n^{k+1} + \nabla_n^{k+2} + \dots \qquad (8.7.4)$$

Met (8.7.4) gaat (8.7.3) over in:

$$\Delta y_n = f_n + \tfrac{1}{2}\nabla f_n + \tfrac{5}{12}\nabla^2 f_n + \tfrac{3}{8}\nabla^3 f_n + \tfrac{251}{720}\nabla^4 f_n + \ldots \qquad (8.7.5)$$

(8.7.5) is de z.g. integratieformule van Adams. De coëfficiënten nemen langzaam af. Tenzij h klein is leveren hoge differenties vaak niet te verwaarlozen bijdragen; ook de in de hogere differenties voorkomende afrondingsfouten dringen tot y_n door: men moet zich ervoor hoeden differenties die niets voorstellen mee te nemen. Met inachtneming van deze voorzorgen is (8.7.5) een bruikbare formule die zeer eenvoudig werkt: men hoeft niets te itereren, maar moet daarentegen een klein interval gebruiken.

Bij gebruik van een sneller convergente centrale reeks kan men een groter interval nemen. B.v. (4.4.4):

$$y_{n+1} = \mu\delta^{-1}_{n+1} - \tfrac{1}{12}\mu\delta_{n+1} + \tfrac{11}{720}\mu\delta^3_{n+1} - \ldots \qquad (8.7.6)$$

Stel dat wij tot en met x_n alles kenden, dus y_n, F_n, $\delta^{-1}_{n+\frac{1}{2}}$ dan kennen wij niet eens de kopterm van (8.7.6):

$$\mu\delta^{-1}_{n+1} = \tfrac{1}{2}(\delta^{-1}_{n+\frac{1}{2}} + \delta^{-1}_{n+3/2})$$

Dit is vrij onbevredigend en wij zullen dan ook trachten op systematische manier te extrapoleren:

Met (4.4.4) vinden wij:

$$y_{n+1} - y_{n-1} = \frac{1}{h}\int_{x_{n-1}}^{x_{n+1}} F\,dx = (\mu\delta^{-1}_{n+1} - \mu\delta^{-1}_{n-1}) - \tfrac{1}{12}(\mu\delta_{n+1} - \mu\delta_{n-1}) + \tfrac{11}{720}(\mu\delta^3_{n+1} - \mu\delta^3_{n-1}) - \ldots \qquad (8.7.7)$$

Wij vinden:

$$\mu\delta^k_{n+1} - \mu\delta^k_{n-1} = 2\,\delta^{k+1}_n + \tfrac{1}{2}\delta^{k+3}_n$$

Met (8.7.8) gaat (8.7.7) over in:

$$y_{n+1} = y_{n-1} + 2F_n + \tfrac{1}{3}\delta^2_n - \tfrac{1}{90}\delta^4_n \;\ldots. \qquad (8.7.8)$$

De gang van zaken is nu deze: bereken met (8.7.8) uit P_n en geschatte differenties een voorlopige y_{n+1}; hieruit P_{n+1} en $\delta^{-1}_{n+3/2}$. Dan met (8.7.6) een betere y_{n+1}. Men moet natuurlijk controleren dat de bij de P_{n+1} behorende differenties y_n zouden kunnen veranderen.
Wij gebruiken dus (8.7.8) als "predictor" en (8.7.6) als "corrector".
De start kunnen wij weer maken met behulp van de Taylor-reeks of met iteratie.

<u>Voorbeeld</u>: $y' = x^2 - y \quad y(0) = 1$

Uit de Taylor-reeks vinden wij $y(x)$ $x = -.3(.1).3$:

x	y	P = hy'				
-.3	1.340141	-.125014				
			7154			
-.2	1.218597	-.117860		1223		
			8377		-117	
-.1	1.104829	-.109483		1106		12
			9483		-105	
0	1.000000	-.100000		1001		9
			10484		- 96	
+.1	.905163	-.089516		905		11
			11389		- 85	
+.2	.821269	-.078127		820		6
			12209		- 79	
+.3	.749182	-.065918		741		9
			12950		- 70	
+.4	.689682	-.052968		671		6
			13621		- 64	
+.5	.643470	-.039347		607		6
			14228		- 58	
+.6	.611190	-.025119		549		7
			14777		- 51	
+.7	.593416	-.010342		498		2
			15275		- 49	
+.8	.590672	+.004933		449		7
			15724		- 42	
+.9	..603433	+.020657		407		
			16131			
+1.0	.632121	+.036788				

Wij hebben hier met Adams geïntegreerd.

Men kan de start ook itereren. Integreert men centraal, dan moet de somfunctie aan de beginvoorwaarde aangepast worden:

$$\mu\delta^{-1}_0 = y_0 + \frac{1}{12}\mu\delta_0 - \frac{11}{720}\mu\delta^3_0 \qquad (8.7.9)$$

Als voorbeeld voor de beginschatting **van y kiezen** wij een constante nl. 1

x	y	δ^{-1}	P	δ	δ^2
		1.336000			
-.3	1.000000		-.091000		
		1.245000		-.005000	
-.2	1.000000		-.096000		+.002000
		1.149000		-.003000	
-.1	1.000000		-.099000		.
		1.050000		-.001000	
0	1.000000		-.100000		.
		0.950000		+.001000	
.1	1.000000		-.099000		.
		0.851000		+.003000	
.2	1.000000		-.096000		.
		0.755000		+.005000	
.3	1.000000		-.091000		
		0.664000			

Toepassing van (8.7.6) geeft nieuwe y en daarmee nieuwe differentie-tabel:

x	y	δ^{-1}	P	δ	δ^2	δ^3
		1.395634				
-.3	1.335325		-.120100			
		1.275534		4367		
-.2	1.217201		-.115733		2399	
		1.159801		6766		-198
-.1	1.104659		-,108967		2201	
		1.050834		8967		-201
0	1.000000		-.100000		2000	
		.950834		10967		-201
.1	.905326		-.089033		1799	
		.861801		12766		-198
.2	.822534		-.076267		1601	
		.785534		14367		
.3	.753325		-.061900			
		.723634				

enz. Na vijf iteraties vindt men geen verandering meer:

x	y	δ^{-1}	P	δ	δ^2	δ^3
		1.403190				
-.3	1.340141		-.125014			
		1.278176		7154		
-.2	1.218597		-.117860		1223	
		1.160316		8377		-117
-.1	1.104829		-.109483		1106	
		1.050833		9483		-105
0	1.000000		-.100000		1001	
		.950833		10484		- 96
.1	.905162		-.089516		905	
		.861317		11389		- 85
.2	.821269		-.078127		820	
		.783190		12209		
.3	.749181		-.065918			
		.717272				

hetgeen vrijwel overeenkomt met de Taylor-reels - start.
De numeriek geïntegreerde waarden van y kloppen goed met de analytische oplossing (die men gemakkelijk vindt)

$$y(x) = 2 - 2x + x^2 - e^{-x}$$

Men vindt hiermee b.v. $y(1) = .6321206$.
Men kan hier, zoals bij tweede orde vergelijkingen, ook zonder somfunctie integreren

$$y_n = \mu\delta^{-1}_n - \frac{1}{12}\mu\delta_n + \frac{11}{720}\mu\delta^3_n + \ldots$$

$$\mu\delta y_n = \mu^2(P_n - \frac{1}{12}\delta^2_n + \frac{11}{720}\delta^4_n \ldots)$$

Nu is $\mu^2 f_n = \frac{1}{4}(f_{n-1} + 2f_n + f_{n+1}) = \frac{1}{4}(4f_n + \delta^2_n) = f_n + \frac{1}{4}\delta^2_n$.

Dus $\mu\delta y_n = P_n + \frac{1}{4}\delta^2_n - \frac{1}{12}\delta^2_n - \frac{1}{48}\delta^4_n + \frac{11}{720}\delta^4_n + \ldots = P_n + \frac{1}{6}\delta^2_n - \frac{1}{180}\delta^4_n$

of

$$y_{n+1} = y_{n-1} + 2P_n + \frac{1}{3}\delta^2_n - \frac{1}{90}\delta^4_n \ldots$$

waarmee wij onze predictor hervonden hebben. Als men dus zonder somfunctie integreert komt dit neer op gebruik van predictor alleen.

Heeft men hogere orde vergelijkingen of tweede orde vergelijkingen waaruit men de 1e afgeleide niet kan verwijderen, dan kan men als in 8.5 te werk gaan en de vergelijking als stelsel eerste orde vergelijkingen schrijven. Deze eerste orde vergelijkingen kan men b.v. volgens 8.7, maar dan simultaan oplossen.

8.8. Meerpuntsrandvoorwaarden

Om de oplossing van een gewone differentiaalvergelijking vast te leggen heeft men randvoorwaarden; in het algemeen evenveel voorwaarden als de orde bedraagt. Bij de eerste orde vergelijking is het eenvoudig: men kan één voorwaarde geven, b.v. de functiewaarde in x_o. Men kan dan in x_o starten en te werk gaan als in 8.7. Bij hogere orde vergelijkingen heeft men de keus tussen de voorwaarden in een punt of in verschillende punten op te leggen. Het eerste is in zoverre eenvoudig, dat men de vergelijking zo nodig als stelsel eerste orde vergelijkingen kan opvatten en van het beginpunt uit simultaan integreren.

Het tweede probleem, dat van meerpuntsvoorwaarden, geeft aanleiding tot essentieel andere technieken. Wij zullen dit aan tweede orde vergelijkingen demonstreren.
Laat voor $y'' = f(x,y,y')$ gegeven zijn de twee functiewaarden y_o in x_o en y_1 in x_1. Men kan nu niet in x_o starten, omdat een gegeven, b.v. de eerste afgeleide, ontbreekt. Wat men doen kan is stellen

$$y_1(x_o) = y_o$$
$$y_1'(x_o) = a_1$$

en integreren tot $x = x_1$: laat men vinden $y_1(x_1) = b_1$.
Laat verder gesteld worden

$$y_2(x_o) = y_o$$
$$y_2'(x_o) = a_2$$

en gevonden worden: $y_2(x_1) = b_2$.

Als men nu de a_1 en a_2 geschikt gekozen heeft, kan men een nieuwe waarde van a bepalen door inverse interpolatie; als men twee waarden a_1 en a_2 genomen heeft, vindt men (lineair):

$$a = a_1 - \frac{a_1 - a_2}{b_1 - b_2}(b_1 - y_1) .$$

Hiermee kan men weer integreren, enz.

Is de differentiaalvergelijking lineair, dan is de zaak eenvoudiger: immers iedere lineaire combinatie van twee oplossingen is een oplossing; wij hebben dan volgens bovenstaande twee oplossingen y_1 en y_2, zodat

$$y_1(x_o) = y_o \qquad y_1(x_1) = b_1$$
$$y_2(x_o) = y_o \qquad y_2(x_1) = b_2$$

en vinden direct

$$y = \frac{b_2 - y_1}{b_2 - b_1} - \frac{b_1 - y_1}{b_2 - b_1} y_2 .$$

Dit loopt mis, als $b_2 \approx b_1$, hetgeen het geval is als de oplossing tengevolge van verandering van beginhelling niet veel verandert.

Een geheel andere methode voor lineaire differentiaalvergelijkingen is de volgende. Men vervangt in de differentiaalvergelijking de afgeleiden door differentie-uitdrukkingen. In ieder punt krijgt men dan een lineaire vergelijking waaraan de functiewaarden van de oplossing moeten voldoen: men krijgt evenveel lineaire vergelijkingen als men onbekende functiewaarden heeft, b.v.

Gevraagd te bepalen y(x), zodat

$$\frac{d^2y}{dx^2} = 2y(y^2-1) \text{ met } y(0) = 0 \text{ en } y(1) = .761590 \qquad (8.8.1)$$

Stel de gekozen intervallengte h, dan is

$$h^2 \frac{d^2y_o}{dx^2} = \delta_o^2 - \frac{1}{12}\delta_o^4 + \frac{1}{90}\delta_o^6 - \dots \qquad (8.8.2)$$

Stel $x_j = jh$ en $y(x_j) = y_j$. (8.8.2) schrijven we als:

$$h^2\frac{d^2y_k}{dx^2} = y_{k-1} - 2y_k + y_{k+1} - \frac{1}{12}\delta_k^4 + \frac{1}{90}\delta_k^6 - \dots \qquad (8.8.3)$$

Dit gesubstitueerd in (8.8.1) geeft:

$$y_{k-1} - 2y_k + y_{k+1} = 2y_k(y_k^2-1)h^2 + \frac{1}{12}\delta_k^4 - \frac{1}{90}\delta_k^6 + \ldots \qquad (8.8.4)$$

met $y_o = 0$ en $y_n = .761590$ als $nh = 1$.

Men zou h zo klein kunnen kiezen, dat de δ^4 en hoger in het rechterlid van (8.8.4) verwaarloosbaar zijn.
Men moet dan oplossen het stelsel:

$$\left.\begin{array}{rl} -2y_1 + y_2 \qquad = & = 2y_1(y_1^2-1)h^2 - y_o \\ y_1 - 2y_2 + y_3 \qquad & = 2y_2(y_2^2-1)h^2 \\ y_2 - 2y_3 + y_4 \qquad & = 2y_3(y_3^2-1)h^2 \\ \ddots \qquad & \\ y_{n-3}-2y_{n-2}+y_{n-1} & = 2y_{n-2}(y_{n-2}^2-1)h^2 \\ y_{n-2}-2y_{n-1} & = 2y_{n-1}(y_{n-1}^2-1)h^2-y_n \end{array}\right\}(8.8.5)$$

(8.8.5) is een niet-lineair stelsel van n-1 vergelijkingen, dat echter reeds zo is opgeschreven dat een oplossingsmethode gesuggereerd wordt: ga (8.8.5) oplossen alsof het een lineair stelsel was en doe dit iteratief, d.w.z. gebruik een oplossing om nieuwe rechterleden te berekenen; los dan weer het stelsel op en bereken rechterleden, enz. Als wij dit doen, is er geen reden meer om de differentiecorrecties niet evenzo te behandelen: bepaal ze als differenties van de vorige oplossing.

De matrix behorende bij het stelsel (8.8.5) is van een zeer bijzonder type:

$$A_n = \{a_{ik}\} \;;\; \begin{array}{ll} a_{ik} = 0 & k \neq i-1, i, i+1 \\ a_{i,i-1} = a_{i,i+1} = 1 & \qquad i,k = 0(1)n-1 \\ a_{ii} \quad = -2 & \end{array}$$

Zoals men gemakkelijk verifieert is de inverse:

$$A_n^{-1} = \frac{-1}{n+1}\{b_{ik}\};\; \begin{array}{ll} b_{ik} = (n-k)(i+1) & \quad i \leqslant k \quad i,k = 0(1)n-1 \\ b_{ik} = b_{ki} \;. & \end{array}$$

Om op ons voorbeeld terug te komen, stellen wij n = 10, dus h = .1 . Als eerste schatting kiezen wij, bij gebrek aan beter een lineaire y:

$$y_i = \frac{i}{10} \cdot 0{,}761590; \quad R = \text{rechterlid}.$$

x	$10^6 y$	δ^4	R	$10^6 y$	δ^4	R	$10^6 y$
0	0			0			0
.1	76159	(0)	-1514	96768	(107)	-1908	99384
.2	152318	0	-2976	192021	107	-3690	196852
.3	228477	0	-4331	284298	456	-5188	290645
.4	304636	0	-5227	372245	-385	-6381	379242
.5	380795	0	-6512	454964	562	-7169	461459
.6	456954	0	-7231	531172	322	-7599	536507
.7	533113	0	-7632	600148	368	-7649	603955
.8	609272	0	-7662	661493	424	-7406	663755
.9	685431	(0)	-768858	715176	(424)	-768542	716148
1.0	761590			761590			761590

De tussen haakjes geplaatste differenties kan men niet berekenen: deze moet men bijschatten; aanvankelijk gaat dat slecht, later kan men deze goed extrapoleren.

Tenslotte vindt men voor de zesde benadering (die nauwelijks meer afwijkt van de vijfde) van y:

x	y	y_{exact}
0	0	0
.1	.09967	.09967
.2	.19737	.19738
.3	.29131	.29131
.4	.37995	.37995
.5	.46212	.46212
.6	.53705	.53705
.7	.60437	.60437
.8	.66404	.66404
.9	.71630	.71630
1.0	.76159	.76159

hetgeen goed overeenstemt met y = th x .

Het alternatief zou zijn integreren van (8.8.1) met geschatte $y'(0)$ tot $x = 1$ en dan door interpolatie nieuwe $y'(0)$ bepalen b.v.

b.v. $y(0) = 0,\ y'(0) = .9$ geeft $y(1) = .67410$

en $y(0) = 0,\ y'(0) = 1.1$ geeft $y(1) = .85357$.

Hieruit zou men door interpolatie vinden $y'(0) = .9975$, waarmee men opnieuw integreert en eventueel door hogere-graads interpolatie weer een nieuwe $y'(0)$ bepaalt.

9. Sommeren van reeksen

Onder sommeren van reeksen verstaan wij hier het numeriek bepalen van de som van een oneindige reeks van termen $f(n)$, waarbij men bij gegeven n, $f(n)$ kan berekenen. Als $f(n)$ voor grote waarde van n sterk, b.v. exponentieel naar nul gaat, dan zal de bepaling van de som geen moeilijkheden opleveren. Convergeert $f(n)$ niet snel naar nul, dan kan het nodig zijn zeer veel termen te bepalen.
Er zijn nu een aantal mogelijkheden. Laat b.v. bekend zijn hoe $f(n)$ zich voor grote waarden van n gedraagt, neem aan als ar^n. In dit geval is het vrij eenvoudig: men bepaalt N zodanig, dat voor $n \geqslant N$, $f(n) = ar^n$ gesteld kan worden en dan

$$\sum_0^{\infty} f(n) = \sum_0^{N-1} f(n) + \sum_N^{\infty} f(n) = \sum_0^{N-1} f(n) + a\sum_N^{\infty} r^n = \sum_0^{N-1} f(n) + \frac{ar^N}{1-r}\,.$$

Een andere mogelijkheid, als $f(n)$ niet alterneert met n, is voor een aantal aequidistante waarden van n, $f(n)$ bepalen; deze waarden interpoleren voor de gehele waarden van n en dan optellen. De interpolatie hoeft men niet echt uit te voeren, het kan volgens de methode van Lubbock.
Stel wij willen bepalen $\sum_{n=a}^{b} f(n)$. Stel $b-a = k.m$, waarbij m zo groot is, dat voor alle waarden van n $f(n)$ door interpolatie te bepalen is uit $f(a)$, $f(a+m)$, $f(a+2m)$ enz. Is $b-a \neq km$, waarbij m zo gekozen is, dat de interpolatie mogelijk is, dan kunnen wij een paar begintermen apart nemen en apart sommeren.

De interpolatie zou kunnen geschieden met een formule van Bessel:

$$f(a+im+n) = f(a+im) + \frac{n}{m}\delta_{\frac{1}{2}+i} + B_2(\tfrac{n}{m})(\delta_i^2+\delta_{1+i}^2) + B_3(\tfrac{n}{m})\delta_{\frac{1}{2}+i}^3 +$$
$$+ B_4(\tfrac{n}{m})(\delta_i^4+\delta_{1+i}^4) + \dots$$

waarbij $0 \leqslant n \leqslant m$ en de differenties de centrale differenties voorstellen bij de basispunten $a+im$ en $a+(i+1)m$.

Om te vinden $\sum_{n=0}^{m} f(a+im+n)$ moeten wij bovenstaande interpolatieformule sommeren; daartoe moeten wij de Besselcoëfficiënten sommeren. De oneven Besselcoëfficiënten zijn antimetrisch t.o.v. het midden van het interval: dus $\sum_{m=0}^{m} B_{2k+1}(\frac{n}{m}) = 0$.

$$\sum_{n=0}^{m} B_2(\tfrac{n}{m}) = \sum_{n=0}^{m} \frac{1}{4}\,\frac{n}{m}(\frac{n}{m}-1) = -\frac{m}{24}\,(1-\frac{1}{m^2}) .$$

Verder vindt men:

$$\sum_{n=0}^{m} B_4(\tfrac{n}{m}) = \frac{11}{1440}\,m(1-\frac{1}{m^2})(1+\frac{1}{11m^2}) .$$

In plaats van te bepalen $\sum_{n=0}^{m} f(a+im+n)$ bepalen wij liever

$$\tfrac{1}{2}f(a+im) + f(a+im+1) + f(a+im+2) + \dots + f(a+(i+1)m-1) + \tfrac{1}{2}f(a+(i+1)m)$$

d.w.z. wij nemen de eerste en laatste term half, opdat wij, als wij de sommen bij elkaar tellen om het totaal te krijgen, niet de eindpunten dubbel krijgen. Voor de sommen van B_{2k} maakt dat niets uit, B_{2k} is immers nul voor $n = 0$ en $n = m$.

Wij krijgen dan voor één interval (van m termen):

$$\tfrac{1}{2}f(a+im) + f(a+im+1) + \dots + f(a+(i+1)m-1) + \tfrac{1}{2}f(a+(i+1)m) =$$

$$\frac{m}{2}f(a+im) + \frac{m}{2}f(a+(i+1)m) - \frac{m}{24}(1-\frac{1}{m^2})(\delta_i^2+\delta_{i+1}^2) +$$
$$+ \frac{11m}{1440}(1-\frac{1}{m^2})(1+\frac{1}{11m^2})(\delta_i^4+\delta_{i+1}^4) + \dots$$

Dit moet gesommeerd worden voor $i = 0(1)k-1$:

$$\sum_{i=0}^{k-1} \left(\tfrac{1}{2}f(a+im) + f(a+im+1) + \ldots + f(a+(i+1)m-1) + \tfrac{1}{2}f(a+(i+1)m)\right) =$$

$$= \frac{m}{2} \sum_{i=0}^{k-1} \left\{(a+im) + f(a+(i+1)m)\right\} - \frac{m}{24}\left(1-\frac{1}{m^2}\right) \sum_{i=0}^{k-1} (\delta_i^2+\delta_{i+1}^2) +$$

$$+ \frac{11m}{1440}\left(1-\frac{1}{m^2}\right)\left(1+\frac{1}{11m^2}\right) \sum_{i=0}^{k-1} (\delta_i^4+\delta_{i+1}^4) + \ldots$$

$$= m\left\{\tfrac{1}{2}f(a) + f(a+m) + \ldots + f(a+(k-1)m) + \tfrac{1}{2}f(a+km)\right\} -$$

$$- \frac{m}{24}\left(1-\frac{1}{m^2}\right) \sum_{i=0}^{k-1} (\delta_i^2+\delta_{i+1}^2) + \frac{11m}{1440}\left(1-\frac{1}{m^2}\right)\left(1+\frac{1}{11m^2}\right) \sum_{i=0}^{k-1} (\delta_i^4+\delta_{i+1}^4) + \ldots$$

Wij hebben nu echter $f(a)$ en $f(b) = f(a+km)$ half geteld; dit moet nog hersteld worden, dus:

$$\sum_a^b f(n) = m \sum_{i=0}^{k-1} f(a+im) - \frac{m}{24}\left(1-\frac{1}{m^2}\right) \sum_{i=0}^{k-1} (\delta_i^2+\delta_{i+1}^2) +$$

$$+ \frac{11m}{1440}\left(1-\frac{1}{m^2}\right)\left(1+\frac{1}{11m^2}\right) \sum_{i=0}^{k-1} (\delta_i^4 + \delta_{i+1}^4) + \ldots$$

Dit is de somformule van Lubbock.

9.2. Alternerende reeksen

De interpolatie gaat dan niet, maar er bestaat een veel betere methode, nl. die van Euler.
Als $f(n)$ alterneert, stellen wij $g_n = (-)^n f(n)$. Dan

$$\sum_0^\infty f(n) = g_o - g_1 + g_2 - g_3 + g_4 - g_5 + \ldots =$$

$$\tfrac{1}{2}g_o - (-\tfrac{1}{2}g_o - \tfrac{1}{2}g_1) + (-\tfrac{1}{2}g_1 + \tfrac{1}{2}g_2) - \ldots =$$

$$\tfrac{1}{2}g_o - \tfrac{1}{2}(\Delta_o - \Delta_1 + \Delta_2 - \Delta_3 + \ldots)$$

Nu hebben wij weer een reeks, in de Δ's, die wij op dezelfde manier behandelen enz.:

$$\sum_0^\infty f(n) = \tfrac{1}{2}g_o - \frac{1}{4}\Delta_o + \frac{1}{8}\Delta_o^2 - \frac{1}{16}\Delta_o^3 + \frac{1}{32}\Delta_o^4 - \ldots$$

Deze tranformatie heet de tranformatie van Euler. N.B. de differenties zijn differenties van g_n, dus van f_n zonder teken! Onder vele omstandigheden convergeert de "geëulerde" reeks sneller dan de oorspronkelijke.

Voorbeeld

$f(n) = (-r)^n$. Eerst differenties maken:

$$\begin{array}{lllll}
1 & & & & \\
 & r-1 & & & \\
r & & (r-1)^2 & & \\
 & r(r-1) & & (r-1)^3 & \dots \\
r^2 & & r(r-1)^2 & & \\
 & r^2(r-1) & & & \\
r^3 & & & &
\end{array}$$

Men ziet de voortzetting, wij vinden:

$$\sum_0^\infty (-r)^n = \frac{1}{2} - \frac{1}{4}(r-1) + \frac{1}{8}(r-1)^2 - \frac{1}{16}(r-1)^3 .$$

Dit is een meetkundige reeks met reden $\frac{r-1}{2}$. De convergentie is verbeterd als $\left|\frac{r-1}{2}\right| < |r|$, dus als $r > \frac{1}{3}$. Is $r < \frac{1}{3}$, dan is de convergentie slechter geworden.

Voorbeeld

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

Stel $x = 1$

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

Zoals men kan nagaan is $\Delta_0^k = \frac{(-)^k}{k+1}$. Eulering geeft dus:

$$\ln 2 = \frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{8}\cdot\frac{1}{3} + \frac{1}{16}\cdot\frac{1}{4} + \frac{1}{32\cdot 5} + \dots$$

Deze reeks is goed bruikbaar; de oorspronkelijke is geheel onbruikbaar.

Bij het toepassen van de Euler-transformatie heeft men de vrijheid te kiezen waar men met "euleren" zal beginnen; dit zal men laten afhangen van de differentietabel: men zoekt de lijn van voorwaartse differenties die de snelste convergentie geeft; de termen tot die lijn sommeert men gewoon.

Begint men b.v. bij de logarithme na de eerste term, dan:

$$\begin{array}{ccccc}
1 & & & & \\
 & -\frac{1}{2} & & & \\
\frac{1}{2} & & \frac{1}{3} & & \\
 & -\frac{1}{6} & & -\frac{1}{4} & \\
\frac{1}{3} & & \frac{1}{12} & & +\frac{1}{5} \\
 & -\frac{1}{12} & & -\frac{1}{20} & \\
\frac{1}{4} & & \frac{1}{30} & & +\frac{1}{30} \\
 & -\frac{1}{20} & & -\frac{1}{60} & \\
\frac{1}{5} & & \frac{1}{60} & & \\
 & -\frac{1}{30} & & & \\
\frac{1}{6} & & & &
\end{array}$$

Wij vinden voor de differenties van de tweede lijn: $\frac{(-)^k}{(k+1)(k+2)}$
en

$$\ln 2 = 1 - \frac{1}{2}\cdot\frac{1}{2} - \frac{1}{4}\cdot\frac{1}{6} - \frac{1}{8}\cdot\frac{1}{12} - \frac{1}{16}\cdot\frac{1}{20} - \frac{1}{32}\cdot\frac{1}{30} - \ldots$$

hetgeen betere convergentie geeft.