TECHNISCHE HOGESCHOOL EINDHOVEN
Afdeling Algemene Wetenschappen
Onderafdeling der Wiskunde

Cursus

Wetenschappelijk Rekenaar A

te Eindhoven

Deel II

J.J.A. Beenakker - P.J. de Doelder - J. Koekoek - C. Ligtmans

Cursus

Wetenschappelijk Rekenaar A

te Eindhoven

Deel II

INHOUD

HOOFDSTUK VII. HET NUMERIEK OPLOSSEN VAN STELSELS LINEAIRE VERGELIJKINGEN

In dit hoofdstuk worden enige methoden behandeld, die gebruikt kunnen worden voor het numeriek oplossen van stelsels lineaire vergelijkingen. Wij onderscheiden directe en indirecte methoden. Onder een directe methode verstaan wij een methode, die in een eindig aantal (van tevoren te bepalen) bewerkingen het juiste antwoord levert, mits van afrondingsfouten geen sprake is. Indirecte methoden zijn iteratieve methoden. Eerst zullen wij een keuze doen uit de vele directe methoden.

1. Eliminatie methode van Gauss

Wij gaan uit van een stelsel van n lineaire vergelijkingen met n onbekenden.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots\ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots\ldots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots\ldots + a_{nn}x_n &= b_n \end{aligned} \qquad (7.1.1)$$

Definieer A = matrix (a_{ij}), $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\underline{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$,

dan kan (7.1.1) worden weergegeven door $A\underline{x} = \underline{b}$.
Veronderstellen wij dat $\det A \neq 0$, dan heeft (7.1.1) precies één oplossing $\underline{x}$. Ga om deze oplossing te bepalen als volgt te werk:
deel de coëfficiënten uit de eerste vergelijking door a_{11} en elimineer daarna met deze vergelijking, (waarin de coëfficiënt van x_1 nu gelijk is aan 1), de x_1 uit de 2^e tot en met de n^e vergelijking.
Wij krijgen dan het aan (7.1.1) equivalente stelsel

$$x_1 + a_{12}^{(1)}x_2 + \ldots\ldots + a_{1n}^{(1)}x_n = b_1^{(1)} \qquad (7.1.2)$$

$$\left.\begin{aligned} a_{22}^{(1)}x_2 + \ldots\ldots + a_{2n}^{(1)}x_n &= b_2^{(1)} \\ \vdots \qquad\qquad & \\ a_{n2}^{(1)}x_2 + \ldots\ldots + a_{nn}^{(1)}x_n &= b_n^{(1)} \end{aligned}\right\} \qquad (7.1.3)$$

Merk op dat, indien $A^{(1)}$ de matrix van het stelsel (7.1.3) voorstelt, geldt $\det A = a_{11} \cdot \det A^{(1)}$.
Behandel nu het stelsel (7.1.3) op dezelfde wijze. Dus deel de eerste vergelijking door $a_{22}^{(1)}$ en elimineer daarna x_2 uit de volgende vergelijkingen. We

zetten dit proces voort en komen uiteindelijk terecht op het stelsel

$$\begin{aligned} x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \ldots\ldots + a_{1n}^{(1)}x_n &= b_1^{(1)} \\ x_2 + a_{23}^{(2)}x_3 + \ldots\ldots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\ x_3 + \ldots\ldots + a_{3n}^{(3)}x_n &= b_3^{(3)} \\ &\vdots \\ x_{n-1} + a_{n-1,n}^{(n-1)}x_n &= b_{n-1}^{(n-1)} \\ x_n &= b_n^{(n)} \end{aligned} \qquad (7.1.4)$$

Uit dit zgn. triangulaire stelsel berekenen wij achtereenvolgens x_n, x_{n-1}, ..., x_1. Bovendien geldt $\det A = a_{11} \cdot a_{22}^{(1)} \cdot a_{33}^{(2)} \cdot \ldots \cdot a_{n,n}^{(n-1)}$.
De coëfficiënten a_{11}, $a_{22}^{(1)}$, $a_{33}^{(2)}$, ..., die bij dit eliminatie proces zo'n speciale rol spelen, worden pivots genoemd.
Het is mogelijk dat tijdens de berekeningen een pivot gelijk aan nul verschijnt. Stel bijv. $a_{11} \neq 0$, $a_{22}^{(1)} = 0$. Door omnummering van de onbekenden $x_2, x_3, \ldots, x_n$ (dus kolommen verwisselen) kunnen we zorgen dat de coëfficiënt links boven in (7.1.3) ongelijk aan nul wordt. De rol van pivot wordt dus nu overgenomen door een andere coëfficiënt uit de eerste vergelijking van (7.1.3). Deze strategie is ook voordelig indien de pivot klein is (dit gebeurt in het algemeen als de pivot verkregen is als het verschil van twee ongeveer even grote getallen (cijferverlies!)). Het delen door zo'n kleine pivot kan grote fouten introduceren. Bij het zoeken naar een geschikte pivot behoeven wij ons niet te beperken tot een enkele vergelijking. Hierop berust de zgn. pivotal condensation. In deze strategie gaan we als volgt te werk:
Door omnummering der onbekenden en door het verwisselen van vergelijkingen kunnen we zorgen dat in het stelsel (7.1.1) de in absolute waarde grootste coëfficiënt in de linker bovenhoek komt te staan. Na de eerste eliminatiestap zoekt men de in absolute waarde grootste coëfficiënt uit de matrix $A^{(1)}$ van het stelsel (7.1.3). Door vernummering van de $(n-1)$ onbekenden en door verwisseling van de $(n-1)$ vergelijkingen uit dit stelsel kan men weer zorgen dat deze coëfficiënt linksboven in (7.1.3) komt te staan etc
Bij de praktische uitvoering van de eliminatie noteert men alleen de coëfficiënten en laat men de verwisseling van rijen en kolommen achterwege. Wij onderstrepen steeds de pivots. Vaak neemt men nog een controlekolom mee: de getallen $c_i = \sum_{j=1}^{n} a_{ij} + b_i$. Voert men op deze kolom dezelfde bewerkingen uit als op de kolom der rechterleden dan blijkt dat de eigenschap dat de termen uit deze kolom de som zijn van de overige termen uit dezelfde rij van het schema, behouden blijft. Dit berust in wezen op het feit dat uit

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad \text{volgt} \quad \sum_{j=1}^{n} a_{ij}(x_j+1) = b_i + \sum_{j=1}^{n} a_{ij} = c_i.$$

Voorbeeld 1

$$1.0134x_1 + 1.9725x_2 + 6.9147x_3 = 2.5824$$
$$-3.0025x_1 + 5.1234x_2 - 1.0925x_3 = 1.4027$$
$$1.8731x_1 + 2.7234x_2 - 1.2576x_3 = 7.2338$$

De coëfficiënten van het stelsel bezitten 5 significante cijfers. Tijdens het rekenen nemen wij een extra cijfer mee. De uitkomsten worden weer afgerond.

Gauss zonder pivotal condensation

x_1	x_2	x_3	b_i	c_i
1.0134	1.9725	6.9147	2.5824	12.4830
-3.0025	5.1234	-1.0925	1.4027	2.4311
1.8731	2.7234	-1.2576	7.2338	10.5727
(1) 1	1.94642	6.82327	2.54825	12.3179
	10.9675	19.3944	9.05382	39.4156
	-0.922439	-14.0383	2.46067	-12.5000
(2)	1	1.76835	0.825514	3.59385
		-12.4071	3.22216	-9.18489
(3)		1	-0.259730	0.740293

De regels (1), (2) en (3) bepalen tezamen het triangulaire stelsel (7.1.4). De oplossingen zijn

$$x_1 = 1.8196 \quad x_2 = 1.2848 \quad x_3 = -0.25970$$

Merk op dat det A = 1.0134 x 10.9675 x -12.4071 = -137.90.

Gauss met pivotal condensation

x_1	x_2	x_3	b_i	c_i
1.0134	1.9725	6.9147	2.5824	12.4830
-3.0025	5.1234	-1.0925	1.4027	2.4311
1.8731	2.7234	-1.2576	7.2338	10.5727
(1) 0.146557	0.285262	1	0.373465	1.80528
-2.84239	5.34505		1.81071	4.40337
2.05741	3.08215		7.70347	12.8430
(2)-0.522974	1		0.333154	0.810180
3.66929			6.67664	10.3459
(3) 1			1.81960	2.81959

$$x_1 = 1.8196 \qquad x_2 = 1.2848 \qquad x_3 = -0.25970.$$

Let op! $6.9147 \times 5.43505 \times 3.66929 = +137.90 = -\det A.$

2. Eliminatie methode van Gauss-Jordan

Hier herleidt men het stelsel $\underline{A}\underline{x} = \underline{b}$ tot een zgn. diagonaal stelsel. Na de eerste eliminatie stap en delen door $a_{22}^{(1)}$ heeft men

$$\begin{aligned}
x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x + \ldots\ldots + a_{1n}^{(1)}x_n &= b_1^{(1)} \\
x_2 + a_{23}^{(2)}x_3 + \ldots\ldots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\
a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 + \ldots\ldots + a_{3n}^{(1)}x_n &= b_3^{(1)} \\
\vdots \qquad\qquad \vdots \qquad\qquad & \vdots \\
a_{n2}^{(1)}x_2 + a_{n3}^{(1)}x_3 + \ldots\ldots + a_{nn}^{(1)}x_n &= b_n^{(1)}
\end{aligned} \tag{7.2.1}$$

Dan elimineert men x_2 niet alleen uit de 3^e t/m de n-de vergelijking maar ook uit de eerste vergelijking. Gaat men zo door dan krijgt men uiteindelijk

$$\begin{aligned}
x_1 &= b_1^{(n)} \\
x_2 &= b_2^{(n)} \\
&\ddots \\
x_n &= b_n^{(n)}
\end{aligned} \tag{7.2.2}$$

De rechterleden van (7.1.6) geven de oplossingen.

3. Eliminatie methode van Crout

Deze methode leidt tot hetzelfde triangulaire stelsel (7.1.4) dat het resultaat was van de Gauss-eliminatie.
Het grote voordeel echter is gelegen in de aanzienlijke beperking van het opschrijven van tussenresultaten; men behoeft slechts de elementen van een hulpmatrix te noteren. Bij stelsels met een groot aantal vergelijkinger geeft deze methode een aanzienlijke tijdsbesparing.

Om het geheel overzichtelijk te maken zullen wij eerst, zonder een bewijs te geven, de gang van zaken schetsen.
We gaan uit van het stelsel $\underline{A}\underline{x} = \underline{b}$

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

en bepalen de hulpmatrix

$$\begin{pmatrix} a'_{11} & a'_{12} & \ldots\ldots & a'_{1n} & n'_1 \\ a'_{21} & a'_{22} & \ldots\ldots & a'_{2n} & b'_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a'_{n1} & a'_{n2} & \ldots\ldots & a'_{nn} & b'_n \end{pmatrix} \qquad (7.3.2)$$

met behulp van de betrekkingen

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

$$a'_{ij} = \frac{1}{a'_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} a'_{ik}\, a'_{kj}\right) \qquad i < j \qquad (7.3.3)$$

$$b'_i = \frac{1}{a'_{ii}} \left(b_i - \sum_{k=1}^{i-1} a'_{ik}\, b'_k\right)$$

met de afspraak $\sum_{k=1}^{o} \ldots = 0$

Daarna vinden wij achtereenvolgens x_n, x_{n-1}, ..., x_1 uit de betrekkingen

$$x_i = b'_i - \sum_{k=i+1}^{n} a'_{ik}\, x_k \qquad i = n, n-1, n-2, \ldots, 1 \qquad (7.3.4)$$

Het bepalen van de elementen van de hulpmatrix (7.3.2) geschiedt op de volgende manier.

Uit (7.3.3) volgt direct

$$a'_{i1} = a_{i1} \qquad i = 1, \ldots, n$$

$$a'_{1j} = \frac{a_{1j}}{a'_{11}} = \frac{a_{1j}}{a_{11}} \qquad j = 2, \ldots, n$$

$$b'_1 = \frac{b_1}{a'_{11}} = \frac{b_1}{a_{11}}$$

Hiermee zijn de elementen uit de eerste kolom en eerste rij van (7.3.2) bepaald. Daarna volgen de overige elementen uit de tweede kolom
$a'_{i2} = a_{i2} - a'_{i1}\, a'_{12} \qquad i \geqslant 2.$
Dan bepalen we nog de onbekende elementen uit de 2e rij.

$$a'_{2j} = \frac{1}{a'_{22}} (a_{2j} - a'_{21}\, a'_{1j}) \qquad j > 2$$

$$b'_2 = \frac{1}{a'_{22}} (b_2 - a'_{21}\, b'_1)$$

Zo vullen wij achtereenvolgens de verschillende kolommen en rijen aan. Om voortdurend een controle te hebben tijdens het rekenen nemen wij een

extra kolom c_i mee: $c_i = \sum_{k=1}^{n} a_{ik} + b_i \qquad i = 1, \ldots, n.$

Deze c_i behandelen we op dezelfde manier als de b_i. Dit komt er op neer, dat we tegelijk met (7.3.1) ook het stelsel $A\underline{y} = \underline{c}$ oplossen.

Wij vinden de y_i uit

$$y_i = c'_i - \sum_{k=i+1}^{n} a'_{ik}\, y_k \qquad i = n, n-1, \ldots, 1. \qquad (7.3.5)$$

Er moet nu gelden (zoals later aangetoond zal worden)

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

(7.3.6) wordt gebruikt als controle op de berekening.

In het geval, dat de matrix A symmetrisch is d.w.z. $a_{ij} = a_{ji}$, kan worden bewezen dat voor de hulpmatrix geldt:

$$a'_{ij} = \frac{a'_{ji}}{a'_{ii}} \qquad i < j.$$ De elementen rechts van de hoofddiagonaal volgen dan direct uit de berekening van de overige elementen.

Een nadeel verbonden aan deze methode is dat pivotal condensation slechts beperkt mogelijk is. Wij zorgen dat wij door omnummeren van de onbekenden en door verwisseling van de vergelijkingen starten met een matrix A waarin $|a_{11}| \geq |a_{ij}|$ $i \geq 1$, $j \geq 1$.

Wij behandelen voorbeeld 1 met deze methode.
Het coëfficiënten schema met hulpkolom is

x_3	x_2	x_1	b_i	c_i
6.9147	1.9725	1.0134	2.5824	12.4830
-1.0925	5.1234	-3.0025	1.4027	2.4311
-1.2576	2.7234	1.8731	7.2338	10.5727

De hulpmatrix met controle kolom wordt

a'_{ij}			b'_i	c'_i
6.9147	0.285262	0.146557	0.373465	1.80528
-1.0925	5.43505	-0.522973	0.333154	0.810180
-1.2576	3.08215	3.66929	1.81960	2.81960

De oplossingen zijn $x_1 = 1.8196$, $x_2 = 1.2848$, $x_3 = -0.25970$. De determinant van de matrix A is op teken na het product van de elementen a'_{11}, a'_{22} en a'_{33}.

Rechtvaardiging van de eliminatie methode van Crout

Wij voeren eerst enige begrippen in.
Een vierkante matrix (a_{ij}) $(ij = 1, 2, \ldots, n)$ waarvoor geldt $a_{ij} = 0$ voor $i < j$ heet een linkertriangulaire matrix (L-matrix). Als $a_{ij} = 0$ voor $i > j$ spreekt men over een rechtertriangulaire matrix (R-matrix).
Wij vermelden enige eigenschappen.
De determinant van een triangulaire matrix is het product van de hoofd diagonaal elementen. Het product van L (resp. R)-matrices is weer een L (resp R)-matrix. De inverse van een triangulaire niet singulaire matrix is triangulair.
We vragen ons nu af of het mogelijk is een vierkante matrix A te schrijven als het product van een L-matrix en een R-matrix : $A = L \cdot R$. Bovendien rijst de vraag of - indien deze "ontbinding" mogelijk is - L en R éénduidig door A zijn bepaald. Dit laatste is inderdaad het geval als wij ons beperken tot R-matrices met hoofddiagonaal elementen gelijk aan 1. Neem

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad L = \begin{pmatrix} a'_{11} & & & \\ a'_{21} & a'_{22} & & \\ \vdots & \vdots & & \\ a'_{n1} & a'_{n2} & & a'_{nn} \end{pmatrix},$$

$$R = \begin{pmatrix} 1 & a'_{12} & a'_{13} & \cdots & a'_{1n} \\ & 1 & a'_{23} & \cdots & a'_{2n} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & 1 \end{pmatrix}$$

Dan volgt met de definitie van matrixvermenigvuldiging uit $A = L \cdot R$

$$\begin{aligned} a_{ij} &= \sum_{k=1}^{j-1} a'_{ik}\, a'_{kj} + a'_{ij} && i \geqslant j \\ a_{ij} &= \sum_{k=1}^{i} a'_{ik}\, a'_{kj} && i < j \end{aligned} \qquad (7.3.7)$$

Dus

$$\begin{aligned} a'_{ij} &= a_{ij} - \sum_{k=1}^{j-1} a'_{ik}\, a'_{kj} && i \geqslant j \\ a'_{ij} &= \frac{1}{a'_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} a'_{ik}\, a'_{kj}\right) && i < j \end{aligned} \qquad (7.3.8)$$

Wij zien dat (7.3.8) identiek is aan (7.3.3). Dit betekent dus dat de matrix (7.3.2) de ontbinding bevat van A in een L- en een R-matrix. Hoe de elementen a'_{ij} uit de betrekkingen (7.3.8) moeten worden bepaald is reeds eerder besproken.
Wij maken nu gebruik van deze triangulaire ontbinding van A om het stelsel vergelijkingen $A\underline{x} = \underline{b}$ op te lossen.
Vermenigvuldiging met de inverse van L geeft $R\underline{x} = L^{-1}\underline{b}$.
Met de notatie $L^{-1}\underline{b} = \underline{b}'$ volgt $\underline{b} = L\underline{b}'$ en dus

$$b_i = \sum_{k=1}^{i} a'_{ik}\, b'_k \quad \text{waaruit volgt}$$

$$b'_i = \frac{1}{a'_{ii}} \left(b_i - \sum_{k=1}^{i-1} a'_{ik}\, b'_k\right) \qquad (7.3.9)$$

Deze betrekking is identiek aan de laatste betrekking van (7.3.3).
De oplossingen van het triangulaire systeem $R\underline{x} = L^{-1}\underline{b} = \underline{b}'$ worden gegeven door (zie (7.3.4))

$$x_i = b'_i - \sum_{k=i+1}^{n} a'_{ik}\, x_k \qquad i = n, n-1, \ldots, 1 \qquad (7.3.10)$$

De betrekking (7.3.6) die als controle wordt gebruikt verifieert men eenvoudig op de volgende manier. Laten $\underline{x}$ en $\underline{y}$ de oplossingen zijn van resp.

$A\underline{x} = \underline{b}$ en $A\underline{y} = \underline{c}$ met $c_i = \sum_{j=1}^{n} a_{ij} + b_i$.

Nu geldt $A(\underline{x} + \underline{1}) = \underline{b} + A\underline{1} = \underline{c}$ waar $\underline{1} = (1, 1, \ldots, 1)$.

Dus $\underline{y} = \underline{x} + \underline{1}$.

Uit $R\underline{y} = L^{-1}\underline{c} = \underline{c}'$ volgt $R(\underline{x} + \underline{1}) = \underline{c}'$ dus $\underline{c}' = \underline{b}' + R\underline{1}$

Uitgeschreven geeft dit

$$c_i' = b_i' + 1 + \sum_{k=i+1}^{n} a_{ik}' \qquad (7.3.11)$$

Tenslotte merken we nog op dat de Gauss eliminatie in wezen berust op een herhaalde vermenigvuldiging van de matrix A met een linker triangulaire matrix ($A = LR$ dus $L^{-1}A = R$).

Tot het oplossen van stelsels lineaire vergelijkingen kan men rekenen het bepalen van de inverse van een gegeven vierkante matrix.
Onder de inverse van een matrix A verstaan wij de matrix B, waarvoor geldt $AB = BA = I$, waarin I de éénheidsmatrix voorstelt.
De n^2 elementen b_{ij} van B moeten bepaald worden uit de n^2 lineaire vergelijkingen:

$$\sum_{j=1}^{n} a_{ij} b_{jk} = \delta_{ik} \qquad \begin{matrix} i = 1, \ldots, n \\ k = 1, \ldots, n \end{matrix} \qquad \text{met } \delta_{ik} = \begin{cases} 0 & \text{als } i \neq k \\ 1 & \text{als } i = k. \end{cases}$$

Neem k vast, dan vinden we de k-de kolom uit B als oplossingen van het stelsel vergelijkingen

$$\begin{aligned}
a_{11}b_{1k} + a_{12}b_{2k} + \ldots + a_{1n}b_{nk} &= 0 \\
a_{21}b_{1k} + a_{22}b_{2k} + \ldots + a_{2n}b_{nk} &= 0 \\
&\vdots \\
a_{k1}b_{1k} + a_{k2}b_{2k} + \ldots + a_{kn}b_{nk} &= 1 \\
&\vdots \\
a_{n1}b_{1k} + a_{n2}b_{2k} + \ldots + a_{nn}b_{nk} &= 0
\end{aligned}$$

De coëfficiëntenmatrix van de linkerkant van (7.3.12) is voor alle stelsels vergelijkingen ter bepaling van de kolommen van B gelijk. We kunnen dus simultaan met één van de besproken eliminatiemethoden de n stelsels oplossen. We gaan uit van het coëfficiëntenschema

$$\begin{array}{cccc|cccc} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \end{array}$$

Aan het slot van de bespreking van enkele eliminatie methoden willen we nog iets zeggen over de gevonden oplossingen.
Veronderstellen we eerst, dat in het stelsel vergelijkingen $A\underline{x} = \underline{b}$ de coëfficiënten a_{ij} en b_i exact zijn (dus geen afgeronde getallen).
Door de noodzakelijke afrondingen tijdens het eliminatieproces, is de gevonden oplossing $\underline{x}'$ niet precies gelijk aan de juiste oplossing $\underline{x}$.
Vullen we de gevonden waarde $\underline{x}'$ in, dan vinden we

$$A\underline{x}' = \underline{b}' \quad \text{met} \quad \underline{x}' = \underline{x} + \delta\underline{x} \quad \text{en} \quad \underline{b}' = \underline{b} + \delta\underline{b}.$$

Dus geldt

$$A(\underline{x} + \delta\underline{x}) = \underline{b} + \delta\underline{b} \quad \text{of} \quad A\delta\underline{x} = \delta\underline{b}. \qquad (7.3.13)$$

Daar het rechterlid van (7.3.13) bekend is, heeft men hier een stelsel lineaire vergelijkingen voor de correctie $\delta\underline{x}$.
In de eliminatie methode van Crout behoeft men ter bepaling van $\delta\underline{x}$ slechts voor één extra kolom (nl. $\delta\underline{b}$) de berekeningen opnieuw uit te voeren.

Het bespreken van de mogelijke fouten in de oplossing, die te wijten zijn aan onnauwkeurigheid in de coëfficiënten (bv. de coëfficiënten zijn afgeronde getallen) laten we achterwege, daar deze materie te gecompliceerd is. We merken alleen nog op, dat soms kleine veranderingen in de waarde der coëfficiënten grote verschuivingen in de oplossingen kunnen veroorzaken. Zulke stelsels vergelijkingen worden "ill-conditioned systems" genoemd.

Iteratieve methoden

Vele stelsels lineaire vergelijkingen, die in de praktijk optreden, kunnen zo worden omgevormd (door omnummering van vergelijkingen en onbekenden), dat de absolute waarde van de coëfficiënt van x_k in de k-de vergelijking groot is in verhouding tot alle andere coëfficiënten uit deze vergelijking. $|a_{kk}| \gg |a_{ki}| \qquad i \neq k.$

Zo'n stelsel vergelijkingen kan men dan vaak door een iteratief proces oplossen.
We behandelen twee methoden.

4. Iteratieve methode van Gauss

Wij schrijven het stelsel $A\underline{x} = \underline{b}$ in de vorm

$$a_{ii}x_i = b_i - \sum_{\substack{k=1 \\ k\neq i}}^{n} a_{ik}x_k \qquad i = 1, \ldots, n \qquad (7.4.1)$$

We kunnen in de rechterleden $\sum_{k=1}^{n} a_{ik}x_k \quad (i \neq k)$ als relatief kleine correcties opvatten en dus als eerste approximatie van de oplossing nemen

$$x_i^{(1)} = \frac{b_i}{a_{ii}} \qquad i = 1, \ldots, n.$$

De volgende approximatie verkrijgt men door in de rechterleden van (7.4.1) voor x_k de eerste approximatie $x_k^{(1)}$ in te vullen.
Men krijgt dan

$$x_i^{(2)} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{k=1 \\ k\neq i}}^{n} a_{ik}x_k^{(1)}\right)$$

Dit proces wordt herhaald tot men de gewenste overeenstemming heeft tussen de nieuw verkregen $x_k^{(\nu+1)}$ en de daarvoor gebruikte $x_k^{(\nu)}$. Het is echter de vraag of dit iteratieproces convergeert.

5. Iteratieve methode van Gauss-Seidel

Wanneer wij tijdens het iteratieproces een onbekende in het rechterlid

steeds vervangen door zijn laatst gevonden approximatie krijgen wij de iteratiemethode van Gauss-Seidel.

Zo krijgt men beginnend met $x_k^{(o)} = 0$, $k=1, \dots, n$

$$
\begin{array}{l}
a_{11}x_1^{(1)} = b_1 \\
a_{22}x_2^{(1)} = b_2 - a_{21}x_1^{(1)} \\
a_{33}x_3^{(1)} = b_3 - a_{31}x_1^{(1)} - a_{32}x_2^{(1)} \\
\vdots \\
\cdots\cdots\cdots\cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots \\
\vdots
\end{array}
$$

en daarna

$$
\begin{array}{l}
a_{11}x_1^{(2)} = b_1 - a_{12}x_2^{(1)} - a_{13}x_3^{(1)} - a_{14}x_4^{(1)} \cdots\cdots \\
a_{22}x_2^{(2)} = b_2 - a_{21}x_1^{(2)} - a_{23}x_3^{(1)} - a_{24}x_4^{(1)} \cdots\cdots \\
a_{33}x_3^{(2)} = b_3 - a_{31}x_1^{(2)} - a_{32}x_2^{(2)} - a_{34}x_4^{(1)} \cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}
$$

Verwacht mag worden dat deze methode, zo ze convergeert, sneller convergeert dan die van Gauss.

Voorbeeld 2

$$
\begin{array}{rcl}
24x_1 - 2x_2 + 2x_3 + x_4 &=& 54 \\
-x_1 + 21x_2 + 2x_3 - x_4 &=& -61 \\
x_1 + 2x_2 + 28x_3 - 2x_4 &=& 28 \\
x_2 - 2x_3 + 20x_4 &=& -45
\end{array}
$$

Wij itereren volgens Gauss-Seidel

x_1	x_2	x_3	x_4	b_i
[24]	-2	2	1	54
-1	[21]	2	-1	-61
1	2	[28]	-2	28
0	1	-2	[20]	-45
2,250	-2,7976	1,1195	-1,9982	
2,00683	-3,01097	1,00067	-1,99938	
1,999004	-3,000082	1,000086	-1,999987	
1,999985	-3,000008	1,000002	-1,999999	
1,999999	-3,000000	1,000000	-2,000000	
2,000000	-3,000000	1,000000	-2,000000	

De convergentie is wegens de relatief grote diagonaalelementen zeer bevredigend.

6. Relaxatie methode

Hierbij schrijft men het stelsel vergelijkingen $A\underline{x} = \underline{b}$ in de vorm

$$R_i = \sum_{k=1}^{n} a_{ik}x_k - b_i \qquad i = 1, 2, \ldots, n \qquad (7.6.1)$$

Substitueert men in (7.6.1) een beginschatting $x_k^{(o)}$ dan zullen de residuën R_i in het algemeen van nul verschillende waarden hebben. Men probeert nu door een herhaalde verbetering van deze eerste schatting te zorgen dat alle residuën nul worden in de gewenste nauwkeurigheid. Men gaat als volgt te werk Men zoekt de in absolute waarde grootste R_i en maakt deze nul door verandering van de $x_k^{(o)}$ met de grootste coëfficiënt. Hierdoor veranderen alle residuën. Van de nieuwe residuën zoekt men weer de grootste enz.

Voorbeeld 3

$$\begin{aligned} 4x_1 + x_2 &= 6 \\ x_1 + 4x_2 - x_3 &= 6 \\ 2x_1 + 2x_2 - 6x_3 &= -12 \end{aligned}$$

$x_1 = 1,\ x_2 = 2,\ x_3 = 3$

$$\begin{aligned} R_1 &= \underline{4x_1} + x_2 - 6 \\ R_2 &= x_1 + \underline{4x_2} - x_3 - 6 \\ R_3 &= 2x_1 + 2x_2 - \underline{6x_3} + 12 \end{aligned}$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	R_1	R_2	R_3
0	0	0	0	-6	-6	<u>12</u>
1	0	0	2	-6	<u>-8</u>	0
2	0	2	2	<u>-4</u>	0	4
3	1	2	2	0	1	<u>6</u>
4	1	2	3	0	0	0

HOOFDSTUK VIII. HET NUMERIEK BEPALEN VAN EIGENWAARDEN EN EIGENVECTOREN VAN MATRICES

1. Algemene theoretische beschouwing

Zij A een lineaire afbeelding van R_n in zich zelf: $\underline{y} = A\underline{x}$. Na keuze van een basis in R_n wordt deze afbeelding gekarakteriseerd door een n- bij n matrix, die wij eveneens met A aangeven.
$\underline{y} = A\underline{x}$ kunnen we dan lezen als een matrixvermenigvuldiging, als we $\underline{x}$ en $\underline{y}$ opvatten als kolomvectoren, dus als matrices met één kolom en n rijen.

Def. Een vector $\underline{x} \neq \underline{0}$ heet een eigenvector van A, als er een getal λ bestaat zodat $A\underline{x} = \lambda\underline{x}$. Het getal λ heet een eigenwaarde van A behorende bij de eigenvector $\underline{x}$.

Om deze eigenvectoren te bepalen moeten we het stelsel lineaire vergelijkingen $A\underline{x} = \lambda\underline{x}$ oplossen. Wij kunnen hiervoor schrijven $(A - \lambda I)\underline{x} = \underline{0}$.
Dit stelsel heeft een van $\underline{0}$ verschillende oplossing, als $\det.(A - \lambda I) = 0$.

Met matrix $A = (a_{ij})$ geeft dit

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots\cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & & a_{nn}-\lambda \end{vmatrix} = 0$$

dus

$$(-1)^n \{\lambda^n - p_1\lambda^{n-1} - p_2\lambda^{n-2} - \cdots\cdots - p_n\} = 0$$

Deze n-de graads vergelijking in λ wordt de karakteristieke vergelijking van de matrix A genoemd. De wortels $\lambda_1, \lambda_2, \ldots, \lambda_n$ van deze vergelijking zijn de eigenwaarden van A.
Nu geldt $p_1 = a_{11} + a_{22} + \cdots\cdots + a_{nn}$ en $p_n = (-1)^{n+1} \det A$.
Uit bekende eigenschappen van nulpunten van polynomen volgt dan

$$\lambda_1 + \lambda_2 + \cdots\cdots + \lambda_n = a_{11} + a_{22} + \cdots\cdots + a_{nn}$$

$$\lambda_1 \cdot \lambda_2 \cdot \cdots\cdots \cdot \lambda_n = \det A.$$

Opmerking: $a_{11} + a_{22} + \cdots\cdots + a_{nn}$ wordt het spoor van de matrix A genoemd, notatie sp(A).

De wortels van de karakteristieke vergelijking kunnen reëel, complex, enkel- of meervoudig zijn. De bijbehorende eigenvectoren worden daarna berekend

uit de vergelijkingen $(A - \lambda I)\underline{x} = \underline{0}$. In het geval dat λ een enkelvoudige wortel is, is de eigenvector $\underline{x}$ bepaald op een multiplicatieve factor na (geen bewijs).
Wanneer λ een p-voudige wortel is, vindt men bij deze eigenwaarde hoogstens p onafhankelijke oplossingen van $(A - \lambda I)\underline{x} = \underline{0}$, dus hoogstens p onafhankelijke eigenvectoren (geen bewijs).

Wij beschouwen in het vervolg slechts zulke matrices A, waarbij een p-voudige wortel van de karakteristieke vergelijking p onafhankelijke eigenvectoren heeft. Dit blijkt bijvoorbeeld het geval te zijn bij symmetrische matrices d.w.z. matrices A waarvoor geldt $A = A^T$. Zie voorbeeld 2.

Naast de matrix A beschouwen we de matrix A^T (de getransponeerde matrix van A). De eigenvectoren en eigenwaarden van A^T vinden wij als de oplossingen van $A^T\underline{y} = \lambda\underline{y}$.
Er geldt, dat de eigenwaarden van A en A^T gelijk zijn, immers
$\det(A - \lambda I) = \det(A - \lambda I)^T = \det(A^T - \lambda I)$, dus de karakteristieke vergelijkingen van A en A^T zijn dezelfde.
Als $\underline{y}$ een eigenvector van A^T is met bijbehorende eigenwaarde λ geldt $A^T\underline{y} = \lambda\underline{y}$ of $\underline{y}^TA = \lambda\underline{y}^T$, waarin $\underline{y}^T$ de getransponeerde van $\underline{y}$ is, dus een rijvector. We noemen $\underline{y}^T$ een eigenrij van A, terwijl de eigenvectoren $\underline{x}$ uit $A\underline{x} = \lambda\underline{x}$ eigenkolommen worden genoemd. Elke matrix van n rijen en n kolommen heeft dus n eigenwaarden, n eigenkolommen en n eigenrijen.
In het geval dat de matrix A symmetrisch is volgt uit $A\underline{x} = \lambda\underline{x}$ dat $\underline{x}^TA = \lambda\underline{x}^T$.
Bij een symmetrische matrix zijn de eigenkolommen en de eigenrijen elkaars getransponeerden.

Voorbeeld 1.

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \\ 1 & 1 & 3 \end{pmatrix} .$$

De karakteristieke vergelijking wordt

$$\begin{vmatrix} 1-\lambda & 2 & 1 \\ 0 & -1-\lambda & 2 \\ 1 & 1 & 3-\lambda \end{vmatrix} = 0 \quad \text{of } -\lambda^3 + 3\lambda^2 + 4\lambda = 0$$
$$\text{met } \lambda_1 = 0,\ \lambda_2 = -1,\ \lambda_3 = 4.$$

De eigenkolommen vindt men uit $A\underline{x} = \lambda\underline{x}$.

$$\lambda_1 = 0 \qquad \begin{aligned} x_1 + 2x_2 + x_3 &= 0 \\ -x_2 + 2x_3 &= 0 \\ x_1 + x_2 + 3x_3 &= 0 \end{aligned} \qquad \underline{x}_1 = \rho \begin{pmatrix} -5 \\ 2 \\ 1 \end{pmatrix}$$

$$\lambda_2 = -1 \qquad \begin{aligned} 2x_1 + 2x_2 + x_3 &= 0 \\ x_3 &= 0 \\ x_1 + x_2 + 4x_3 &= 0 \end{aligned} \qquad \underline{x}_2 = \sigma \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

$$\lambda_3 = 4 \qquad \begin{aligned} -3x_1 + 3x_2 + x_3 &= 0 \\ -5x_2 + 2x_3 &= 0 \\ x_1 + x_2 - x_3 &= 0 \end{aligned} \qquad \underline{x}_3 = \tau \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}$$

De eigenrijen van A vindt men als de getransponeerden van de eigenkolommen van A^T : $A^T y = \lambda y$.

$$A^T = \begin{pmatrix} 1 & 0 & 1 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$$

Wij vinden

$$\lambda_1 = 0 \qquad \underline{y}_1^T = \alpha(1, 1, -1)$$
$$\lambda_2 = -1 \qquad \underline{y}_2^T = \beta(2, 7, -4)$$
$$\lambda_3 = 4 \qquad \underline{y}_3^T = \gamma(1, 1, 3)$$

Stelling

Laten $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ eigenvectoren van A zijn met bijbehorende eigenwaarden $\lambda_1, \lambda_2, \ldots, \lambda_m$. Stel bovendien, dat de λ_i twee aan twee verschillend zijn, dan zijn de eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ lineair onafhankelijk.

Bewijs

Stel $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ zijn niet lineair onafhankelijk, dan spannen zij een lineaire deelruimte van R_n op met een dimensie k, met $k < m$.
Door omnummering der $\underline{x}_i$ en λ_i kan men bereiken, dat als basis voor deze deelruimte $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ kan worden genomen. De overige $\underline{x}_i$ $(i > k)$ zijn lineair in deze basiselementen uit te drukken, bijv. $\underline{x}_m = \sum_{i=1}^{k} \alpha_i \underline{x}_i$ met bijv. $\alpha_h \neq 0$.

Dan geldt

$$\sum_{i=1}^{k} \lambda_m \alpha_i \underline{x}_i = \lambda_m \underline{x}_m = A\underline{x}_m = A\left(\sum_{i=1}^{k} \alpha_i \underline{x}_i\right) =$$

$$\sum_{i=1}^{k} \alpha_i (A\,\underline{x}_i) = \sum_{i=1}^{k} \alpha_i \lambda_i \underline{x}_i$$

dus $$\sum_{i=1}^{k} (\lambda_m - \lambda_i)\alpha_i \underline{x}_i = \underline{0}.$$

Daar $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ lineair onafhankelijk zijn volgt hieruit $(\lambda_m - \lambda_i)\,\alpha_i = 0$ voor $i = 1, 2, \ldots, k$, dus $(\lambda_m - \lambda_h)\,\alpha_h = 0$ en daar $\alpha_h \neq 0$ geldt $\lambda_m = \lambda_h$, wat in strijd is met de veronderstelling, dat de λ_i *twee aan twee verschillend zijn.*

Conclusie

Elke matrix A bezit (met de reeds vermalde restrictie) n onafhankelijke eigenvectoren (eigenkolommen en eigenrijen). Deze eigenvectoren kan men als een basis van R_n nemen.

Zij $\underline{x}_k$ een eigenkolom van A met bijbehorende eigenwaarde λ_k en zij $\underline{y}_j^T$ een eigenrij van A met bijbehorende eigenwaarde λ_j, $\lambda_j \neq \lambda_k$, dan geldt

$$\lambda_j\, \underline{y}_k^T \underline{x}_j = \underline{y}_k^T (\lambda_j\, \underline{x}_j) = \underline{y}_k^T (A\,\underline{x}_j) = (\underline{y}_k^T A)\,\underline{x}_j = \lambda_k\, \underline{y}_k^T \underline{x}_j,$$

dus $$(\lambda_j - \lambda_k)\, \underline{y}_k^T \underline{x}_j = 0 \quad \text{of} \quad \underline{y}_k^T \underline{x}_j = 0.$$

Het inwendig product van de twee vectoren $\underline{y}_k^T$ en $\underline{x}_j$ is dus 0, wat betekent dat $\underline{y}_k^T$ en $\underline{x}_j$ orthogonaal zijn.
Controleer dit resultaat bij voorbeeld 1.

Stelling

Als de karakteristieke vergelijking van de matrix A n verschillende wortels heeft, bestaan er matrices X en Y zodanig, dat YAX = D, waarin D een diagonaalmatrix is met in de hoofddiagonaal de eigenwaarde λ_i.

Bewijs

Neem voor X de matrix, waarin de kolommen de eigenkolommen van A zijn en voor Y de matrix met in de rijen de eigenrijen van A. Normeer bovendien de

$\underline{x}_i$ en de $\underline{y}_i^T$ zó, dat $\underline{y}_i^T \underline{x}_i = 1$.

Dan geldt voor het element d_{ij} van D

$$d_{ij} = \underline{y}_i^T (A \underline{x}_j) = \lambda_j \underline{y}_i^T \underline{x}_j, \quad \text{dus} \quad \begin{matrix} d_{ij} = 0 \\ d_{ii} = \lambda_i \end{matrix} \qquad \text{voor } i \neq j.$$

In het geval van voorbeeld 1 krijgen wij voor de genormeerde eigenvectoren

$$\underline{x}_1 = \begin{pmatrix} -5 \\ 2 \\ 1 \end{pmatrix}, \quad \underline{y}_1^T = (-\tfrac{1}{4}, -\tfrac{1}{4}, \tfrac{1}{4})$$

$$\underline{x}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \underline{y}_2^T = (-\tfrac{2}{5}, -\tfrac{7}{5}, \tfrac{4}{5})$$

$$\underline{x}_3 = \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}, \quad \underline{y}_3^T = (\tfrac{1}{20}, \tfrac{1}{20}, \tfrac{3}{20})$$

de matrices X en Y worden dan

$$X = \begin{pmatrix} -5 & 1 & 3 \\ 2 & -1 & 2 \\ 1 & 0 & 5 \end{pmatrix} \qquad Y = \begin{pmatrix} -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{2}{5} & -\frac{7}{5} & \frac{4}{5} \\ \frac{1}{20} & \frac{1}{20} & \frac{3}{20} \end{pmatrix}$$

Controleer, dat $YAX = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$

2. Symmetrische matrices

Een symmetrische matrix A is een matrix waarvoor geldt $A^T = A$. We zagen reeds, dat de eigenrijen van een symmetrische matrix de getransponeerden van de eigenkolommen zijn.
We geven nu nog enkele eigenschappen.

Stelling

De eigenwaarden van een symmetrische matrix zijn reëel.

Bewijs

Stel λ is een complexe eigenwaarde van de reële symmetrische matrix A en $\underline{x}$ de bijbehorende complexe eigenvector, dan volgt uit $A\underline{x} = \lambda\underline{x}$ dat $\overline{A\,\underline{x}} = \overline{\lambda\,\underline{x}}$ dus $A\,\overline{\underline{x}} = \overline{\lambda}\,\overline{\underline{x}}$, hierin zijn $\overline{\lambda}$ en $\overline{\underline{x}}$ de toegevoegde complexe waarden van λ resp. $\underline{x}$.

Dan geldt $\lambda\,\overline{\underline{x}^T}\cdot\underline{x} = \overline{\underline{x}^T}\cdot A\,\underline{x} = \overline{\lambda}\,\overline{\underline{x}^T}\cdot\underline{x}$, dus $(\lambda - \overline{\lambda})\,\overline{\underline{x}^T}\cdot\underline{x} = 0$. Daar $\overline{\underline{x}^T}\cdot\underline{x} \neq 0$ volgt $\lambda = \overline{\lambda}$ of λ is reëel.

Stelling

Eigenvectoren van een symmetrische matrix behorende bij verschillende eigenwaarden zijn orthogonaal : $\underline{x}_k^T\,\underline{x}_j = 0$, $\lambda_k \neq \lambda_j$. (voor bewijs zie eerder)

Stelling

Heeft men in de karakteristieke vergelijking van de symmetrische matrix A een p-voudige wortel λ, dan kan men bij deze λ p onafhankelijke eigenvectoren $\underline{x}_1, \ldots, \underline{x}_p$ vinden (geen bewijs).

Wij gaan nu deze p eigenvectoren $\underline{x}_1, \ldots, \underline{x}_p$ (behorende bij een p-voudige wortel λ) orthogonaliseren.

Ter vereenvoudiging beperken wij ons tot het geval $p = 2$.

Als $\underline{x}_1$ en $\underline{x}_2$ onafhankelijke eigenvectoren van A zijn bij dezelfde λ, dan is ook $\alpha\,\underline{x}_1 + \underline{x}_2$ een eigenvector van A met dezelfde eigenwaarde λ, want

$$A(\alpha\underline{x}_1 + \underline{x}_2) = \alpha A\underline{x}_1 + A\underline{x}_2 = \alpha\lambda\underline{x}_1 + \lambda\underline{x}_2 = \lambda(\alpha\underline{x}_1 + \underline{x}_2).$$

Bepaal nu α zó, dat $\underline{x}_1^T(\alpha\underline{x}_1 + \underline{x}_2) = 0$, neem dus

$$\alpha = -\frac{\underline{x}_1^T\cdot\underline{x}_2}{\underline{x}_1^T\cdot\underline{x}_1}.$$ De eigenvectoren $\underline{x}_1$ en $\underline{x}_2' = \alpha\underline{x}_1 + \underline{x}_2$ zijn nu orthogonaal.

Conclusie

Een symmetrische matrix A bezit n lineair onafhankelijke eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$, die twee aan twee loodrecht op elkaar staan.

Stelling

Bij een symmetrische matrix A bestaat er een matrix X zodanig, dat $X^TAX = D$, $X^T = X^{-1}$, waar D een diagonaalmatrix is met in de hoofddiagonaal de eigenwaarden van A.

Bewijs

A heeft eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$, die loodrecht op elkaar staan. Normeer bovendien de $\underline{x}_i$ zó, dat $\underline{x}_i^T \underline{x}_i = 1$. Neem voor X de matrix, waarvan de kolommen de eigenkolommen $\underline{x}_i$ zijn. Dan geldt voor het element d_{ij} van D

$$d_{ij} = \underline{x}_i^T A \underline{x}_j = \lambda_j \underline{x}_i^T \underline{x}_j \qquad \text{dus } d_{ij} = 0 \quad \text{voor } i \neq j$$
$$d_{ii} = \lambda_i$$

Bovendien geldt $X^T X = I$, want $\underline{x}_k^T \underline{x}_j = 0$ als $k \neq j$
$= 1$ als $k = j$.

De matrix X is dus orthogonaal.

Voorbeeld 2

$$A = \begin{pmatrix} 1 & -4 & 8 \\ -4 & 7 & 4 \\ 8 & 4 & 1 \end{pmatrix}, \quad A^T = A$$

De karakteristieke vergelijking wordt

$$\begin{vmatrix} 1-\lambda & -4 & 8 \\ -4 & 7-\lambda & 4 \\ 8 & 4 & 1-\lambda \end{vmatrix} = 0 \quad \text{of} \quad \lambda^3 - 9\lambda^2 - 81\lambda + 729 = 0$$
$$(\lambda + 9)(\lambda - 9)^2 = 0$$
$$\lambda_1 = -9 \qquad \lambda_2 = \lambda_3 = 9.$$

De eigenvectoren vinden we uit

$\lambda_1 = -9$

$$\begin{aligned} 10x_1 - 4x_2 + 8x_3 &= 0 \\ -4x_1 + 16x_2 + 4x_3 &= 0 \\ 3x_1 + 4x_2 + 10x_3 &= 0 \end{aligned}$$

$\underline{x}_1 = \sigma \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}$; kies $\sigma = 1$

$\lambda_2 = \lambda_3 = 9$

$$\begin{aligned} -8x_1 - 4x_2 + 8x_3 &= 0 \\ -4x_1 - 2x_2 + 4x_3 &= 0 \\ 8x_1 + 4x_2 - 8x_3 &= 0 \end{aligned}$$

of $2x_1 + x_2 - 2x_3 = 0$

Hiervan zijn de oplossingen $\underline{x} = \lambda \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$

Kies $\underline{x}_2 = \begin{pmatrix}1\\0\\1\end{pmatrix}$ en bepaal daarna μ zó, dat $\underline{x}_3 = \begin{pmatrix}1\\0\\1\end{pmatrix} + \mu\begin{pmatrix}0\\2\\1\end{pmatrix}$ loodrecht staat op $\underline{x}_2$. We vinden $\mu = -2$ en $\underline{x}_3 = \begin{pmatrix}1\\-4\\-1\end{pmatrix}$

Hierna normeren we zó, dat $\underline{x}_k^T \underline{x}_k = 1$.

We krijgen

$$\underline{x}_1 = \frac{1}{3}\begin{pmatrix}2\\1\\-2\end{pmatrix}, \quad \underline{x}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}, \quad \underline{x}_3 = \frac{1}{\sqrt{18}}\begin{pmatrix}1\\-4\\-1\end{pmatrix}$$

De matrix X wordt $\begin{pmatrix} \frac{1}{3} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} \\ \frac{1}{3} & 0 & -\frac{4}{\sqrt{18}} \\ -\frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \end{pmatrix}$

Ga na, dat voldaan is aan

$$X^T A X = \begin{pmatrix} -9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

3. <u>Het numeriek bepalen van eigenwaarden en eigenvectoren</u>

Om de eigenwaarden te bepalen kan men de karakteristieke vergelijking opstellen en daarna oplossen. De bijbehorende eigenvectoren kunnen dan worden verkregen door stelsels lineaire vergelijkingen op te lossen. Bij omvangrijke matrices geeft dit zeer veel rekenwerk. Bovendien is het in de praktijk meestal niet te doen om alle eigenwaarden, maar slechts om enkele speciale eigenwaarden. Met behulp van iteratieve methoden kan men vaak deze speciale eigenwaarden met bijbehorende eigenvectoren op een handige manier berekenen.

4. <u>Iteratieve methode ter bepaling van de grootste eigenwaarde van een matrix met de daarbij behorende eigenvector</u>.

Zij A een reële matrix van de orde n met eigenwaarden $\lambda_1, \ldots, \lambda_n$ en

eigenvectoren $\underline{x}_1, \ldots, \underline{x}_n$. We veronderstellen dat λ_1 reëel is en dat $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \ldots$. Uitgaande van een willekeurige vector $\underline{v} = \alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \ldots + \alpha_n\underline{x}_n$ met $\alpha_1 \neq 0$ bepalen we achtereenvolgens de vectoren $A\underline{v}$, $A^2\underline{v}$, $A^3\underline{v}$, ..., $A^m\underline{v}$.
De vectoren $A^m\underline{v}$ worden (door A) geïtereerde vectoren genoemd.
Daar $A^m\underline{x}_i = \lambda_i^m \underline{x}_i$ geldt

$$A^m\underline{v} = \alpha_1\lambda_1^m \underline{x}_1 + \alpha_2\lambda_2^m \underline{x}_2 + \ldots + \alpha_n\lambda_n^m \underline{x}_n$$

of

$$A^m\underline{v} = \alpha_1\lambda_1^m \left[\underline{x}_1 + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^m \underline{x}_2 + \ldots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^m \underline{x}_n\right] \qquad (8.4.1)$$

Omdat $|\lambda_1| > |\lambda_i| \qquad i = 2, \ldots, n$ volgt hieruit

$$\lim_{m\to\infty} \frac{A^m\underline{v}}{\alpha_1\lambda_1^m} = \underline{x}_1$$

Voor grote waarde van m geldt dan binnen de zelf te kiezen nauwkeurigheidsgrenzen, dat $A^m\underline{v} = \alpha_1\lambda_1^m \underline{x}_1$.
Dus $A^m\underline{v}$ is een eigenvector behorende bij de eigenwaarde λ_1

De eigenwaarde λ_1 vinden we uit $\dfrac{(A^{m+1}\underline{v})i}{(A^m\underline{v})i}$, waarin $(A^m\underline{v})i$ de i-de component van $A^m\underline{v}$ is.

Is λ_1 in absolute waarde >1 of <1, dan worden de componenten van de geïtereerde vectoren spoedig òf onhandelbaar groot òf onhandelbaar klein zie (8.4.1). Men kan dit verhelpen door na iedere iteratie op een passende wijze te normeren. Het beste kan men één bepaalde component van de geïtereerde vector steeds weer gelijk aan 1 maken door na iedere iteratie alle componenten door deze component te delen. Wij nemen hiervoor bij voorkeur de component die in absolute waarde de grootste is. D.w.z. in plaats van de rij $A\underline{v}$, $A^2\underline{v}$, $A^3\underline{v}$, ... beschouwen we de rij $\underline{v}_1, \underline{v}_2, \underline{v}_3, \ldots$ gedefinieerd door

$$\underline{v}_1 = \frac{A\underline{v}}{(A\underline{v}, \underline{e}_k)} \; ; \quad \underline{v}_{m+1} = \frac{A\underline{v}_m}{(A\underline{v}_m, \underline{e}_k)} \qquad m = 1, 2, \ldots \qquad (8.4.2)$$

($\underline{e}_k$ is de k-de éénheidsvector, de k-de component van $\underline{v}_{m+1}$ is dus 1)

Uit (8.4.2) volgt

$$\underline{v}_m = \frac{A^m\underline{v}}{(A^m\underline{v},\underline{e}_k)} = \frac{\underline{x}_1 + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^m \underline{x}_2 + \cdots\cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^m \underline{x}_n}{(\underline{x}_1,\ \underline{e}_k) + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^m (\underline{x}_2,\underline{e}_k) + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^m (\underline{x}_n,\underline{e}_k)} \qquad (8.4.3)$$

Dus geldt $\lim_{m\to\infty} \underline{v}_m = \frac{\underline{x}_1}{(\underline{x}_1,\underline{e}_k)}$ mits $(\underline{x}_1,\ \underline{e}_k) \neq 0$.

Hiermee is de eigenvector behorende bij de grootste eigenwaarde bepaald.
De eigenwaarde λ_1 halen we uit de betrekking

$$(A\underline{v}_m,\ \underline{e}_k) = \lambda_1 \frac{1 + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^{m+1} \frac{(\underline{x}_2,\underline{e}_k)}{(\underline{x}_1,\underline{e}_k)} + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^{m+1} \frac{(\underline{x}_n,\underline{e}_k)}{(\underline{x}_1,\underline{e}_k)}}{1 + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^{m} \frac{(\underline{x}_2,\underline{e}_k)}{(\underline{x}_1,\underline{e}_k)} + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^{m} \frac{(\underline{x}_n,\underline{e}_k)}{(\underline{x}_1,\underline{e}_k)}} \qquad (8.4.4)$$

Er geldt $\lim_{m\to\infty} (A\underline{v}_m,\ \underline{e}_k) = \lambda_1$.

Deze speciale normering heeft dus het voordeel dat we gedurende het rekenen de convergentie tot stand zien komen.
De snelheid waarmee het iteratieproces convergeert is afhankelijk van het quotiënt $\left|\frac{\lambda_2}{\lambda_1}\right|$ (de zgn. convergentie factor). Naarmate de convergentie factor kleiner is verloopt het proces sneller. Een slechte convergentie treedt op als $\left|\frac{\lambda_2}{\lambda_1}\right| \approx 1$.

Pas voor zeer grote waarde van m vinden wij een redelijke benadering van de eigenwaarde λ_1 en de eigenvector $\underline{x}_1$. Hoe we in dit geval te werk gaan zullen we later behandelen.
Een slechte convergentie kan ook optreden als voor de startvector $\underline{v}$ een ongelukkige keuze is gedaan.
Stel $\underline{v} = \alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \cdots + \alpha_n\underline{x}_n$ met $|\alpha_1|$ klein (in het geval dat A symmetrisch is betekent dit $\underline{v}$ nagenoeg loodrecht op $\underline{x}_1$)

$$A^m\underline{v} = \alpha_1\lambda_1^m\underline{x}_1 + \alpha_2\lambda_2^m\underline{x}_2 + \cdots + \alpha_n\lambda_n^m\underline{x}_n$$

Pas bij zeer grote waarden van m zal $\alpha_1\lambda_1^m\underline{x}_1$ gaan overheersen en de convergentie zal zeer langzaam tot stand komen. In het geval dat $\alpha_1 = 0$ zal op de

duur door afrondingsfouten een component in de $\underline{x}_1$ richting ontstaan en dus weer hetzelfde gelden.

Voorbeeld 3

$$A = \begin{pmatrix} 855 & 861 & 80 & -133 \\ 2268 & 2619 & 244 & -396 \\ 1344 & 1512 & 147 & -231 \\ 1428 & 1848 & 168 & -273 \end{pmatrix}$$

Wij willen λ_1 en $\underline{x}_1$ bepalen in zes significante cijfers.
De convergentie is uitstekend omdat hier $\frac{\lambda_1}{\lambda_2} \approx 40$

Als controle nemen wij een extra rij mee, die men verkrijgt door de elementen uit iedere kolom op te tellen. Deze rij behandelen we op dezelfde manier als de overigen.
In het begin kunnen we volstaan met slechts weinig decimalen te rekenen.

	x_1	x_2	x_3	x_4	
	855	861	80	-133	
	2268	2619	244	-396	
	1344	1512	147	-231	
	1428	1848	168	-273	
	5895	6840	639	-1033	
$\underline{v}$	1	1	1	1	
$A\underline{v}$	1663	4735	2772	3171	12341
$\underline{v}_1$	0.35	1	0.58	0.67	2.60
$A\underline{v}_1$	1118	3289	1913	2262	8582
$\underline{v}_2$	0.340	1	0.582	0.688	2.609
$A\underline{v}_2$	1106.76	3259.68	1895.59	2243.47	8505.49
$\underline{v}_3$	0.33953	1	0.58153	0.68825	2.60930
$A\underline{v}_3$	1106.283	3258.400	1894.827	2242.654	8502.165
$\underline{v}_4$	0.339517	1	0.581521	0.688268	2.609307
$A\underline{v}_4$	1106.269	3258.362	1894.805	2242.629	
$\underline{v}_5$	0.339517	1	0.581521	0.688268	

Dus $\lambda_1 = 3258.362$ en $\underline{x}_1 = \begin{pmatrix} 0.339517 \\ 1 \\ 0.581521 \\ 0.688268 \end{pmatrix}$

Is de matrix A symmetrisch dan kan men de convergentie van de benaderingen voor de eigenwaarde (niet van die voor de eigenvector) als volgt versnellen. Wij mogen in dit geval veronderstellen dat de eigenvectoren een orthonormaal stelsel vormen. (d.w.z. $(\underline{x}_i, \underline{x}_j) = \delta_{ij}$)
Uit (8.4.3) volgt dan

$$\frac{(A\underline{v}_m, \underline{v}_m)}{(\underline{v}_m, \underline{v}_m)} = \lambda_1 \frac{1 + \left(\frac{\alpha_2}{\alpha_1}\right)^2 \left(\frac{\lambda_2}{\lambda_1}\right)^{2m+1} + \ldots + \left(\frac{\alpha_n}{\alpha_1}\right)^2 \left(\frac{\lambda_n}{\lambda_1}\right)^{2m+1}}{1 + \left(\frac{\alpha_2}{\alpha_1}\right)^2 \left(\frac{\lambda_2}{\lambda_1}\right)^{2m} + \ldots + \left(\frac{\alpha_n}{\alpha_1}\right)^2 \left(\frac{\lambda_n}{\lambda_1}\right)^{2m}} \qquad (8.4.5)$$

De convergentiefactor is dan dus $\left|\frac{\lambda_2}{\lambda_1}\right|^2$

Uitdrukking (8.4.5) wordt het Rayleigh quotiënt genoemd.

Een andere normering tijdens het iteratie proces krijgt men als volgt: na iedere stap reduceert men de lengte van de geïtereerde vector tot 1. Wij beschouwen dus de rij $\underline{v}_1$, $\underline{v}_2$, ... gedefiniëerd door

$$\underline{v}_1 = \frac{A\underline{v}}{\sqrt{(A\underline{v}, A\underline{v})}}, \quad \underline{v}_{m+1} = \frac{A\underline{v}_m}{\sqrt{(A\underline{v}_m, A\underline{v}_m)}} \qquad m = 1, 2, \ldots.$$

Hieruit volgt

$$\underline{v}_m = \frac{A^m\underline{v}}{\sqrt{(A^m\underline{v}, A^m\underline{v})}}$$

Men kan eenvoudig nagaan dat $\lim_{m\to\infty} (A\underline{v}_m, A\underline{v}_m) = \lambda_1^2$, terwijl $\underline{v}_{2m}$ convergeert naar een eigenvector behorende bij λ_1. Het teken van λ_1 vinden we door voor "grote"waarden van m de vectoren $\underline{v}_m$ en $A\underline{v}_m$ te beschouwen, want dan geldt $A\underline{v}_m \approx \lambda_1 \underline{v}_m$ waaruit direct volgt of wij het + dan wel het - teken moeten nemen.

Iteratieve methoden ter bepaling van eigenwaarden en eigenvectoren in het geval, dat $|\lambda_1| = |\lambda_2| > |\lambda_3| \ldots$.
We onderscheiden hier drie mogelijkheden

$$\lambda_1 = \lambda_2,\ \lambda_1 = -\lambda_2 \qquad \text{en} \qquad \lambda_1 = \overline{\lambda}_2$$

5. $\lambda_1 = \lambda_2$ $\qquad |\lambda_1| > |\lambda_3|$ .... λ_1 reëel.
Zijn $\underline{x}_1$ en $\underline{x}_2$ twee onafhankelijke eigenvectoren behorende bij de tweevoudige eigenwaarde λ_1, dan is ook $\alpha_1\underline{x}_1 + \alpha_2\underline{x}_2$ een eigenvector behorende bij λ_1, want $A(\alpha_1\underline{x}_1 + \alpha_2\underline{x}_2) = \lambda_1(\alpha_1\underline{x}_1 + \alpha_2\underline{x}_2)$.
We gaan weer uit van een willekeurige startvector $\underline{v}$

$$\underline{v} = \alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \alpha_3\underline{x}_3 + \ldots + \alpha_n\underline{x}_n$$

$$A^m\underline{v} = \lambda_1^m\left[\alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \left(\frac{\lambda_3}{\lambda_1}\right)^m \alpha_3\underline{x}_3 + \ldots + \left(\frac{\lambda_n}{\lambda_1}\right)^m \alpha_n\underline{x}_n\right] \qquad (8.5.1)$$

Voor m groot genoeg geldt

$$A^m\underline{v} = \lambda_1^m\left[\alpha_1\underline{x}_1 + \alpha_2\underline{x}_2\right] \quad , \quad \frac{(A^{m+1}\underline{v}, \underline{e}_k)}{(A^m\underline{v}, \underline{e}_k)} = \lambda_1 \qquad (8.5.2)$$

Door uit te gaan van een andere startvector vinden wij een tweede eigenvector.

6. $\lambda_1 = -\lambda_2$, $|\lambda_1| > |\lambda_3|$ ..... λ_1 is reëel.
Het is mogelijk voor dit geval een speciale methode te geven ter bepaling van de eigenwaarden en eigenvectoren. Wij doen dit echter niet wegens het risico dat men deze methode ook zal hanteren als $\lambda_1 \approx -\lambda_2$ (bijv. $\lambda_1 = 40$, $\lambda_2 = -39.9$). We prefereren daarom de methode die besproken zal worden in paragraaf 8.

7. $|\lambda_1| = |\lambda_2|$, λ_1 en λ_2 zijn complex. $|\lambda_1| > |\lambda_3| > \ldots\ldots$.
De karakteristieke vergelijking bezit reële coëfficiënten, dus $\lambda_2 = \overline{\lambda}_1$.
Als $\underline{x}_1$ een (complexe) eigenvector behorende bij λ_1 is, dan is $\underline{x}_2 = \overline{\underline{x}}_1$ een

eigenvector behorende $\lambda_2 = \bar{\lambda}_1$, want

$$A\,\underline{x}_2 = A\,\bar{\underline{x}}_1 = \overline{A\,\underline{x}_1} = \bar{\lambda}_1\bar{\underline{x}}_1 = \lambda_2\underline{x}_2.$$

We itereren weer uitgaande van een willekeurige reële vector $\underline{v}$

$$\underline{v} = \alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \alpha_3\underline{x}_3 + \ldots + \alpha_n\underline{x}_n$$

Omdat de vector reële componenten heeft (op de natuurlijke basis) geldt $\bar{\underline{v}} = \underline{v}$ dus

$$\underline{0} = \bar{\underline{v}} - \underline{v} = (\bar{\alpha}_1\bar{\underline{x}}_1 + \bar{\alpha}_2\bar{\underline{x}}_2 + \ldots + \bar{\alpha}_n\bar{\underline{x}}_n) -$$

$$- (\alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \ldots + \alpha_n\underline{x}_n) = (\bar{\alpha}_2 - \alpha_1)\,\underline{x}_1 + (\bar{\alpha}_1 - \alpha_2)\,\underline{x}_2 + \ldots,$$

waaruit volgt $\alpha_2 = \bar{\alpha}_1$.

Na iteratie krijgen we

$$A^m\underline{v} = \lambda_1^m\alpha_1\underline{x}_1 + \bar{\lambda}_1^m\bar{\alpha}_1\bar{\underline{x}}_1 + \lambda_1^m\left[\alpha_3\underline{x}_3\left(\frac{\lambda_3}{\lambda_1}\right)^m + \ldots + \alpha_n\underline{x}_n\left(\frac{\lambda_n}{\lambda_1}\right)^m\right] \tag{8.7.1}$$

Noemen wij de vector binnen de vierkante haken $\underline{R}_m$ en bedenken we, dat $z + \bar{z} = 2R_e z$, dan krijgen we

$$A^m\underline{v} = 2R_e(\lambda_1^m\alpha_1\underline{x}_1) + \lambda_1^m\underline{R}_m\,, \qquad \lim_{m\to\infty}\underline{R}_m = \underline{0}. \tag{8.7.2}$$

Stel $\lambda_1 = re^{i\varphi}$, dan geldt

$$\frac{A^m\underline{v}}{r^m} = 2R_e(e^{im\varphi}\alpha_1\underline{x}_1) + e^{im\varphi}\underline{R}_m\,, \qquad \lim_{m\to\infty} e^{im\varphi}\underline{R}_m = \underline{0} \tag{8.7.3}$$

De vector $\alpha_1\underline{x}_1$ zal hierin onder invloed van de multiplicatieve factor $e^{im\varphi}$ bij elke iteratie over een hoek φ gedraaid worden, zodat $R_e(e^{im\varphi}\alpha_1\underline{x}_1)$ en dus $\frac{A^m\underline{v}}{r^m}$ een zeer onregelmatig verloop zal hebben (met o.a. tekenwisseling der componenten) en niet zal convergeren.

Het onregelmatig verloop van de geïtereerde vectoren geeft ons juist de aanwijzing, dat we met complexe eigenwaarden te doen hebben. Toch kunnen we met behulp van deze iteraties de complexe eigenwaarden en eigenvectoren berekenen.

Wij beschouwen voor voldoende grote m drie opéénvolgende iteraties $A^m\underline{v}$, $A^{m+1}\underline{v}$ en $A^{m+2}\underline{v}$ en nemen aan, dat wij in (8.7.2) de staart $\lambda_1^m\underline{R}_m$ mogen verwaarlozen t.o.v. $2R_e(\lambda_1^m\alpha_1\underline{x}_1)$.
Deze veronderstelling is volledig speculatief, daar wij immers niet kunnen zien, of dit inderdaad voor deze waarde van m geldt.
(Denk aan het ronddraaien van de vector $\lambda_1^m\alpha_1\underline{x}_1$).

Dan geldt

$$\begin{aligned} A^m\underline{v} &= c_1\underline{x}_1 + c_2\underline{x}_2 \\ A^{m+1}\underline{v} &= \lambda_1 c_1\underline{x}_1 + \lambda_2 c_2\underline{x}_2 \\ A^{m+2}\underline{v} &= \lambda_1^2 c_1\underline{x}_1 + \lambda_2^2 c_2\underline{x}_2 \\ c_2 &= \bar{c}_1,\ \underline{x}_2 = \bar{\underline{x}}_1,\ \lambda_2 = \bar{\lambda}_1 \end{aligned} \qquad (8.7.4)$$

De drie vectoren $A^m\underline{v}$, $A^{m+1}\underline{v}$ en $A^{m+2}\underline{v}$ liggen dus in het vlak opgespannen door de vectoren $\underline{x}_1$ en $\underline{x}_2$ en zijn dus lineair afhankelijk. Er bestaat een lineaire betrekking van de vorm

$$A^{m+2}\underline{v} + a_1A^{m+1}\underline{v} + a_oA^m\underline{v} = \underline{0}. \qquad (8.7.5)$$

Hieruit volgt met gebruikmaking van (8.7.4)

$$(\lambda_1^2 + a_1\lambda_1 + a_o)\, c_1\underline{x}_1 + (\lambda_2^2 + a_1\lambda_2 + a_o)\, c_2\underline{x}_2 = \underline{0}.$$

Daar $\underline{x}_1$ en $\underline{x}_2$ lineair onafhankelijk zijn, volgt, dat λ_1 en λ_2 wortels zijn van

$$\lambda^2 + a_1\lambda + a_o = 0.$$

De bijbehorende eigenvectoren $\underline{z}_1$ en $\underline{z}_2$ halen we uit (8.7.4).
We vinden

$$\begin{aligned} \underline{z}_1 &= A^{m+1}\underline{v} - \lambda_2A^m\underline{v} \\ \underline{z}_2 &= A^{m+1}\underline{v} - \lambda_1A^m\underline{v} \end{aligned}$$

of $$\underline{z}_{1,2} = (A^{m+1}\underline{v} - \alpha A^m\underline{v}) \mp i\beta A^m\underline{v} \qquad \lambda_{1,2} = \alpha \mp i\beta \qquad (8.7.6)$$

Met deze gevonden schattingen voor de eigenvectoren kan men niet verder itereren. Men kan met de oude $\underline{v}_m$ opnieuw enkele iteraties uitvoeren en weer de vierkantsvergelijking oplossen en zien of de gevonden eigenwaarden en eigenvectoren veranderd zijn of niet.
We kunnen ook controleren of binnen de vereiste nauwkeurigheid voldaan is aan $A\,\underline{x} = \lambda\underline{x}$.

Wanneer men tijdens het iteratieproces op de bekende wijze normeert dan beschouwt men in de plaats van het drietal vectoren $A^m\underline{v}$, $A^{m+1}\underline{v}$ en $A^{m+2}\underline{v}$ de vectoren $\underline{v}_m$, $A\underline{v}_m$ en $(A\underline{v}_m, \underline{e}_k)$. $A\underline{v}_{m+1}$
De coëfficiënten a_o en a_1 van de vierkantsvergelijking worden berekend uit de betrekking

$$(A\underline{v}_m, \underline{e}_k)\, A\underline{v}_{m+1} + a_1 A\underline{v}_m + a_o \underline{v}_m = \underline{0}$$

Waarna men voor de eigenvectoren $\underline{z}_1$, $\underline{z}_2$ behorend bij λ_1, λ_2 vindt

$$\underline{z}_{1,2} = (A\underline{v}_m - \alpha \underline{v}_m) \pm i\beta \underline{v}_m \quad , \qquad \lambda_{1,2} = \alpha \pm i\beta$$

Voorbeeld 4

$$A = \begin{pmatrix} 15 & 32 & 8 & -48 \\ 80 & 79 & 40 & -80 \\ -96 & -112 & -49 & 128 \\ 64 & 48 & 32 & -33 \end{pmatrix}$$

	x_1	x_2	x_3	x_4	
	15	32	8	-48	
	80	79	40	-80	
	-96	-112	-49	128	
	64	48	32	-33	
	63	47	31	-33	Controle
$\underline{v}_1$	1	1	1	1	
$A\underline{v}_1$	7	119	-129	111	108
$\underline{v}_2$	-0.05	-0.92	1	-0.86	-0.83
$A\underline{v}_2$	19.09	32.12	-51.24	13.02	12.99
$\underline{v}_3$	-0.373	-0.627	1	-0.254	-0.254
$A\underline{v}_3$	-5.467	-19.05	24.52	-13.59	-13.587
$\underline{v}_4$	-0.223	-0.777	1	-0.554	-0.554
$A\underline{v}_4$	6.3830	5.0970	-11.480	-1.2860	-1.2860
$\underline{v}_5$	-0.5560	0.4440	1	0.1120	0.1120
$A\underline{v}_5$	-19.92400	-48.51600	68.44000	-28.59200	-28.59200
$\underline{v}_6$	-0.291116	-0.708884	1	-0.417767	-0.417767
$A\underline{v}_6$	1.001788	-5.869756	4.867968	-6.871545	-6.871545
$\underline{v}_7$	0.205792	-1.205792	1	-1.411584	-1.411584
$A\underline{v}_7$	40.25757	74.13251	-114.3901	33.87494	33.87494

Beschouw $\underline{v}_6$, $A\,\underline{v}_6$ en $(A\,\underline{v}_6, \underline{e}_3) \cdot A\,\underline{v}_7$ en onderzoek deze drie vectoren op afhankelijkheid.

$$\underline{v}_6 = (-0.291116,\ -0.708884,\ 1\ ,\ -0.417767)$$
$$A\,\underline{v}_6 = (1.001788\ ,\ -5.869756,\ 4.867968\ ,\ -6.871545)$$
$$(A\underline{v}_6, \underline{e}_3) \cdot A\,\underline{v}_7 = (195.9726\ ,\ 360.8747,\ -556.8472,\ 164.9021)$$

De coëfficiënten a_o en a_1 in de afhankelijkheidsbetrekking

$$(A\,\underline{v}_6, \underline{e}_3) \cdot A\,\underline{v}_7 + a_1 A\,\underline{v}_6 + a_o \underline{v}_6 = \underline{0} \qquad (8.7.7)$$

vinden we uit de vergelijkingen

$$195.9726 + 1.001788a_1 - 0.291116a_o = 0,$$
$$-556.8472 + 4.867968a_1 + a_o = 0.$$

Wij vinden $a_1 = -14.00018$
$a_o = 624.9996$

Hierna controleren we of bij deze waarden van a_o en a_1 ook door de twee andere kolommen op bevredigende wijze aan (8.7.7) is voldaan.

Het rechterlid van (8.7.7) wordt na invullen van de gevonden waarden a_o en a_1
voor de 2e kolom 0.000124
voor de 4e kolom 0.000759

Het resultaat is inderdaad bevredigend. Wanneer dit niet het geval zou zijn geweest, moesten we nog enkele iteraties uitvoeren en opnieuw op afhankelijkheid toetsen.

De vierkantsvergelijking voor de waarden λ_1 en λ_2 wordt

$$\lambda^2 - 14.00018 + 624.9996 = 0$$
$$\lambda_{1,2} = 7.00009 \pm 23.99999\ i \text{ (de exacte waarden zijn } 7 \pm 24i).$$

De eigenvectoren $\underline{z}_1$ en $\underline{z}_2$ vinden we uit

$$\underline{z}_{1,2} = (A\,\underline{v}_6 - \alpha \underline{v}_6) \pm i\beta\,\underline{v}_6 \quad \text{met } \lambda_{1,2} = \alpha \pm i\beta.$$

$$\underline{z}_{1,2} = \begin{pmatrix} 3.03960 \\ -0.90757 \\ -2.13212 \\ -3.94718 \end{pmatrix} \pm i \begin{pmatrix} -6.98678 \\ -17.0132 \\ 24.0000 \\ -10.0264 \end{pmatrix}$$

Ingevuld in $A\,\underline{x} = \lambda\,\underline{x}$ vinden we

$$A.\,\underline{x} = \begin{pmatrix} 188.959 \\ 401.960 \\ -590.919 \\ 213.000 \end{pmatrix} \pm i \begin{pmatrix} 24.0427 \\ -140.874 \\ 116.832 \\ -164.917 \end{pmatrix}$$

$$\lambda\,\underline{x} = \begin{pmatrix} 188.960 \\ 401.964 \\ -590.925 \\ 213.003 \end{pmatrix} \pm i \begin{pmatrix} 24.0423 \\ -140.876 \\ 116.831 \\ -164.918 \end{pmatrix}$$

Er is een verschil van maximaal zes eenheden in het zesde cijfer.

8. Bijna gelijke eigenwaarden $|\lambda_1| \approx |\lambda_2| > |\lambda_3| \ldots$

λ_1 en λ_2 zijn reëel. Hieronder valt dus het geval $\lambda_1 = -\lambda_2$.

Dat wij in dit geval verkeren merken we aan de zeer slechte convergentie van het iteratieproces.
Uitgaande van de vector $\underline{v} = \alpha_1\underline{x}_1 + \alpha_2\underline{x}_2 + \alpha_3\underline{x}_3 + \ldots + \alpha_n\underline{x}_n$
vinden we voor de m-de iteratie

$$A^m\underline{v} = \lambda_1^m\alpha_1\underline{x}_1 + \lambda_2^m\alpha_2\underline{x}_2 + \lambda_1^m\left[\left(\frac{\lambda_3}{\lambda_1}\right)^m \alpha_3\underline{x}_3 + \ldots + \left(\frac{\lambda_n}{\lambda_1}\right)^m \alpha_n\underline{x}_n\right] \qquad (8.8.1)$$

Voor voldoende grote waarde van m geldt

$$A^m\underline{v} = \lambda_1^m\alpha_1\underline{x}_1 + \lambda_2^m\alpha_2\underline{x}_2.$$

Wij passen nu de zojuist onder 7. beschreven methode toe.
Beschouw drie opeenvolgende iteraties.

$$\begin{aligned} A^m\underline{v} &= c_1\underline{x}_1 + c_2\underline{x}_2 \quad , \\ A^{m+1}\underline{v} &= \lambda_1 c_1\underline{x}_1 + \lambda_2 c_2\underline{x}_2 \quad , \\ A^{m+2}\underline{v} &= \lambda_1^2 c_1\underline{x}_1 + \lambda_2^2 c_2\underline{x}_2 \quad . \end{aligned}$$

De afhankelijkheidsbetrekking $A^{m+2}\underline{v} + a_1A^{m+1}\underline{v} + a_oA^m\underline{v} = \underline{0}$ levert ons de vierkantsvergelijking $\lambda^2 + a_1\lambda + a_o = 0$ voor de eigenwaarden λ_1 en λ_2.

De bijbehorende eigenvectoren $\underline{z}_1$ en $\underline{z}_2$ zijn

$$\lambda_1 : \underline{z}_1 = A^{m+1}\underline{v} - \lambda_2 A^m \underline{v} \quad ,$$
$$\lambda_2 : \underline{z}_2 = A^{m+1}\underline{v} - \lambda_1 A^m \underline{v} \quad .$$

Met de gevonden schatting voor $\underline{x}_1$ itereren we verder en bepalen $\underline{x}_1$ en λ_1 in de vereiste nauwkeurigheid.
Hoe we aan een betere schatting voor λ_2 en $\underline{x}_2$ komen wordt later behandeld.

Wanneer men tijdens het iteratieproces op de bekende wijze normeert werkt men met de betrekkingen

$$(A\,\underline{v}_m, \underline{e}_k) A\,\underline{v}_{m+1} + a_1 A\,\underline{v}_m + a_o\,\underline{v}_m = \underline{0}$$

$$\lambda^2 + a_1\lambda + a_o = 0 \qquad (8.8.2)$$

$$\underline{z}_{1,2} = A\,\underline{v}_m - \lambda_{2,1}\,\underline{v}_m$$

Voorbeeld 5

$$A = \begin{pmatrix} 9.08 & -1.03 & 1.04 \\ 3.99 & 4.10 & -3.92 \\ 3.97 & -3.92 & 5.94 \end{pmatrix}$$

	x_1	x_2	x_3	
	9.08 3.99 3.97	-1.03 4.10 -3.92	1.04 -3.92 5.94	
	17.04	-0.85	3.06	
$\underline{v}$	1	1	1	
$A\,\underline{v}$	9.09	4.17	5.99	19.25
$\underline{v}_1$	1.00	0.46	0.66	2.12
$A\,\underline{v}_1$	9.293	3.289	6.087	18.669
$\underline{v}_2$	1.000	0.354	0.655	2.009
$A\,\underline{v}_2$	9.3966	2.8738	6.4730	18.7434
$\underline{v}_3$	1.0000	0.3058	0.6889	1.9947
$A\,\underline{v}_3$	9.48148	2.54329	6.86333	18.88810
$\underline{v}_4$	1.00000	0.26824	0.72387	1.99211
$A\,\underline{v}_4$	9.55654	2.25221	7.21829	19.02704
$\underline{v}_5$	1.00000	0.23567	0.75532	1.99099
$A\,\underline{v}_5$	9.622793	1.995393	7.532774	19.150960
$\underline{v}_6$	1.000000	0.207361	0.782805	1.990166
$A\,\underline{v}_6$	9.680535	1.771585	7.807007	19.259127
$\underline{v}_7$	1.000000	0.183005	0.806464	1.989469
$A\,\underline{v}_7$	9.730227	1.578982	8.043017	19.352226

De convergentie is zeer slecht, wat wijst op ongeveer gelijke eigenwaarden. Onderzoek de vectoren $\underline{v}_6$, $A\,\underline{v}_6$ en $(A\,\underline{v}_6, \underline{e}_1) \cdot A\,\underline{v}_7$ op afhankelijkheid.

$$\underline{v}_6 = (1.000000\ ,\ 0.207361\ ,\ 0.782805\ )$$
$$A\,\underline{v}_6 = (9.680535\ ,\ 1.771585\ ,\ 7.807007\ )$$
$$(A\underline{v}_6, \underline{e}_1) \cdot A\,\underline{v}_7 = (94.193803,\ 15.285391,\ 77.860708)$$

Bereken a_1 en a_0 met behulp van de 1e en 3e kolom.
We vinden

$$a_1 = -18.011701$$
$$a_0 = 80.169099$$

Controle in kolom 2 geeft 0.000076.

De vierkantsvergelijking wordt $\lambda^2 - 18.011701\lambda + 80.169099 = 0$

$$\lambda_1 = 9.972447$$
$$\lambda_2 = 8.038253$$
$$\underline{z}_1 = (1.642282\ ,\ 0.104765,\ 1.514622)$$
$$\underline{z}_2 = (-0.292912,\ -0.296519,\ -0.000257)$$

We normeren $\underline{z}_1$ tot $\underline{v} = (1, 0.063792, 0.922267)$ en itereren weer verder

$\underline{v}$	1.000000	0.063792	0.922267
$A\,\underline{v}$	9.973452	0.636261	9.198201
$\underline{v}_1$	1.000000	0.063795	0.922269
$A\,\underline{v}_1$	9.973451	0.636265	9.198201
$\underline{v}_2$	1.000000	0.063796	0.922269
$A\,\underline{v}_2$	9.973450	0.636269	9.198198
$\underline{v}_3$	1.000000	0.063796	0.922268
$A\,\underline{v}_3$	9.973449	0.636273	9.198192
$\underline{v}_4$	1.000000	0.063797	0.922268
$A\,\underline{v}_4$	9.973448	0.636277	9.198188
$\underline{v}_5$	1.000000	0.063797	0.922268

Hieruit volgt $\lambda_1 = 9.973448$ met bijbehorende eigenvector $\underline{z}_1 =$ (1.000000, 0.063797, 0.922268).

9. Extrapolatiemethode van Aitken

Wij hebben gezien dat in het geval van bijna gelijke eigenwaarden het iteratieproces zeer langzaam convergeert. We kunnen de convergentie versnellen door middel van de extrapolatie methode van Aitken, ook wel δ^2-proces genoemd. We gaan uit van een iteratieproces dat ons een rij schattingen y_n geeft voor een grootheid y en we veronderstellen dat geldt

$$y_{n+1} - y \approx c(y_n - y) \quad n = 0, 1, \ldots, \; |c| < 1$$

We spreken in dit geval over een proces van de eerste orde of lineair proces met convergentiefactor c ($|c| < 1$)
Bij iedere nieuwe schatting wordt de afwijking van y ongeveer met een constante factor c vermenigvuldigd.

Nu geldt

$$y_n - y \approx c^n v \quad \text{met} \quad v = y_o - y$$

Men neemt nu drie opeenvolgende schattingen

$$y_o = y + v, \quad y_1 \approx y + cv, \quad y_2 \approx y + c^2 v$$

Hieruit kunnen we een betere schatting voor y afleiden.
We werken met behulp van differenties.
In voorwaartse differenties uitgedrukt hebben we

$$\Delta_o = y_1 - y_o \approx (c-1)v \qquad \Delta_o^2 = (y_2 - y_1) - (y_1 - y_o) \approx (c-1)^2 v$$

dus

$$\frac{(\Delta_o)^2}{\Delta_o^2} \approx v \quad \text{of} \quad y \approx y_o - \frac{(\Delta_o)^2}{\Delta_o^2}$$

Werkend met centrale en achterwaartse differenties vinden we

$$y \approx y_1 - \frac{\delta\frac{1}{2} \cdot \delta\frac{3}{2}}{\delta_1^2}, \quad y \approx y_2 - \frac{(\nabla_2)^2}{\nabla_2^2}, \quad y \approx y_1 - \frac{\Delta_o \nabla_2}{\delta_1^2}$$

We passen nu deze methode toe op de componenten van drie opeenvolgende vectoren $\underline{v}_m$, $\underline{v}_{m+1}$ en $\underline{v}_{m+2}$ uit voorbeeld 5.

$\underline{v}_6$	1.000000	0.207361	0.782805
$\underline{v}_7$	1.000000	0.183005	0.806464
$\underline{v}_8$	1.000000	0.162276	0.826601

We extrapoleren en vinden voor de nieuwe componenten

$$0.207361 - \frac{(-0.024356)^2}{0.003627} = 0.043806$$

$$0.782805 - \frac{(0.023569)^2}{-0.003522} = 0.941734 .$$

Hiermee verder itererend vinden we

	x_1	x_2	x_3
$\underline{v}$	1.000000	0.043806	0.941734
A $\underline{v}$	10.014283	0.478007	9.392180
$\underline{v}_1$	1.000000	0.047733	0.937878
A $\underline{v}_1$	10.006228	0.509224	9.353882
$\underline{v}_2$	1.000000	0.050891	0.934806

Opnieuw extrapolerend met $\underline{v}$, $\underline{v}_1$ en $\underline{v}_2$ vinden we

$$0.043806 - \frac{(0.003927)^2}{-0.000769} = 0.063860$$

$$0.941734 - \frac{(-0.003856)^2}{0.000784} = 0.922769$$

Hiermee is de juiste waarde al veel dichter benaderd. Bij verder itereren en extrapoleren wordt de noemer Δ_0^2 steeds kleiner en de waarde van het quotiënt $\frac{(\Delta_0)^2}{\Delta_0^2}$ steeds onnauwkeuriger. Extrapoleren heeft dan geen zin meer.

10. <u>Bepaling van de overige eigenwaarden van symmetrische matrices met de deflatiemethode van Hotelling.</u>

Zij A een symmetrische matrix van de orde n met de (reële) eigenwaarden $\lambda_1, \lambda_2, \ldots, \lambda_n$ en de bijbehorende eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. We veronderstellen, dat de grootste eigenwaarde λ_1 en de bijbehorende eigenvector $\underline{x}_1$ reeds bepaald zijn. Verder nemen wij aan dat de eigenvectoren een ortho-

normaal stelsel vormen ($\underline{x}_i^T \underline{x}_j = \delta_{ij}$). We voeren nu een nieuwe matrix A_1 in door te definiëren

$$A_1 = A - \lambda_1 \underline{x}_1 \underline{x}_1^T . \qquad (8.10.1)$$

Bewering: deze matrix van de orde n bezit de eigenwaarden 0, λ_2, λ_3, ..., λ_n en de eigenvectoren $\underline{x}_1$, $\underline{x}_2$, $\underline{x}_3$, ..., $\underline{x}_n$.

Bewijs

$$A_1 \underline{x}_1 = A \underline{x}_1 - \lambda_1 \underline{x}_1 \underline{x}_1^T \underline{x}_1 = \lambda_1 \underline{x}_1 - \lambda_1 \underline{x}_1 = \underline{0}$$

$$A_1 \underline{x}_k = A \underline{x}_k - \lambda_1 \underline{x}_1 \underline{x}_1^T \underline{x}_k = \lambda_k \underline{x}_k \qquad k = 2, \ldots, n$$

want $\underline{x}_1^T \underline{x}_k = 0$, als $k \neq 1$.

Om nu λ_2 en $\underline{x}_2$ te bepalen, gaan we uit van een willekeurige startvector $\underline{v}$ en itereren deze met behulp van de matrix A_1 op de gewone manier

$$\underline{v} = \alpha_1 \underline{x}_1 + \alpha_2 \underline{x}_2 + \ldots + \alpha_n \underline{x}_n$$

$$A_1^m \underline{v} = \lambda_2^m \left[\alpha_2 \underline{x}_2 + \alpha_3 \left(\frac{\lambda_3}{\lambda_2}\right)^m \underline{x}_3 + \ldots + \alpha_n \left(\frac{\lambda_n}{\lambda_2}\right)^m \underline{x}_n\right] \qquad (8.10.2)$$

Door het itereren is de component in de $\underline{x}_1$ richting verdwenen en kunnen we op de bekende manier λ_2 en $\underline{x}_2$ bepalen.

Voorbeeld 6

Van de matrix A zijn de grootste eigenwaarde λ_1 en de bijbehorende eigenvector $\underline{x}_1$ gegeven

$$A = \begin{pmatrix} 8.79 & 4.38 & 7.01 \\ 4.38 & 4.04 & 2.63 \\ 7.01 & 2.63 & 9.75 \end{pmatrix} \qquad \lambda_1 = 18.03566$$

$$\underline{x}_1 = \begin{pmatrix} 1.000000 \\ 0.501874 \\ 1.005342 \end{pmatrix}$$

Gevraagd worden λ_2 en $\underline{x}_2$.

Allereerst normeren we $\underline{x}_1$ zó, dat $\underline{x}_1^T \underline{x}_1 = 1$.
We vinden

$$\underline{x}_1 = \begin{pmatrix} 0.664809 \\ 0.333650 \\ 0.668361 \end{pmatrix}$$

Daarna berekenen we de matrix $\underline{x}_1\underline{x}_1^T$

$$\underline{x}_1\underline{x}_1^T = \begin{pmatrix} 0.441971 & 0.221814 & 0.444332 \\ 0.221814 & 0.111322 & 0.222999 \\ 0.444322 & 0.222999 & 0.446706 \end{pmatrix}$$

Daarna bepalen we $A_1 = A - \lambda_1 \underline{x}_1\underline{x}_1^T$

$$A_1 = \begin{pmatrix} 0.818761 & 0.379438 & -1.003821 \\ 0.379438 & 2.032234 & -1.003821 \\ -1.003821 & -1.391934 & 1.693362 \end{pmatrix}$$

startend met de vector $\underline{v} = (1,1,1)$ krijgen we

	x_1	x_2	x_3
$\underline{v}$	1	1	1
$A_1\underline{v}$	0.194	1.020	-0.702
$\underline{v}_1$	0.190	1.000	-0.688
$A_1\underline{v}_1$	1.226	3.062	-2.748
$\underline{v}_2$	0.400	1.000	-0.897
$A_1\underline{v}_2$	1.607	3.433	-3.312
$\underline{v}_3$	0.468	1.000	-0.965
$A_1\underline{v}_3$	1.7313	3.5530	-3.4958
$\underline{v}_4$	0.4873	1.0000	-0.9839
$A_1\underline{v}_4$	1.7661	3.5867	-3.5472
$\underline{v}_5$	0.4924	1.0000	-0.9890
$A_1\underline{v}_5$	1.7754	3.5957	-3.5610
$\underline{v}_6$	0.4938	1.0000	-0.9903
$A_1\underline{v}_6$	1.77783	3.59803	-3.56456
$\underline{v}_7$	0.49411	1.00000	-0.99070
$A_1\underline{v}_7$	1.778481	3.598707	-3.565546
$\underline{v}_8$	0.494200	1.000000	-0.990785
$A_1\underline{v}_8$	1.778640	3.598860	-3.565780
$\underline{v}_9$	0.494223	1.000000	-0.990808
$A_1\underline{v}_9$	1.778682	3.598900	-3.565842
$\underline{v}_{10}$	0.494229	1.000000	-0.990814
$A_1\underline{v}_{10}$	1.778693	3.598911	-3.565858
$\underline{v}_{11}$	0.494231	1.000000	-0.990816
$A_1\underline{v}_{11}$	1.778697	3.598915	-3.565864
$\underline{v}_{12}$	0.494231	1.000000	-0.990816

Dus $\lambda_2 = 3.59892$ $\qquad \underline{x}_2 = \begin{pmatrix} 0.49423 \\ 1.00000 \\ -0.99082 \end{pmatrix}$

We ronden af op 5 decimalen, omdat in de matrix A_1 de zesde decimaal door afrondingen is ontstaan. Ingevuld in de matrix A vinden we

$$A\,\underline{x}_2 = \begin{pmatrix} 1.77867 \\ 3.59889 \\ -3.56590 \end{pmatrix} \qquad \text{en} \qquad \lambda_2\underline{x}_2 = \begin{pmatrix} 1.77870 \\ 3.59892 \\ -3.56586 \end{pmatrix}$$

We kunnen dit deflatie proces van Hotelling voortzetten en zo successievelijk de overige eigenwaarden en eigenvectoren berekenen. Wij krijgen dus de rij matrices

$$A_k = A_{k-1} - \lambda_k \underline{x}_k \underline{x}_k^T \qquad k = 1, 2, \ldots, n \qquad (A_o = A)$$

De matrix A_k heeft de eigenwaarden $0, 0, \ldots, 0, \lambda_{k+1}, \ldots, \lambda_n$ en de eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$.
Hieruit volgt dat A_n geen van nul verschillende eigenwaarden bezit en dus de nulmatrix is. Dit betekent dat A geschreven kan worden als de som van n matrices van de rang 1

$$A = \lambda_1 \underline{x}_1 \underline{x}_1^T + \lambda_2 \underline{x}_2 \underline{x}_2^T + \ldots + \lambda_n \underline{x}_n \underline{x}_n^T$$

11. <u>Het berekenen van overige eigenwaarden van niet symmetrische matrices met de deflatie methode van Hotelling.</u>

We onderscheiden hier twee gevallen:

a) λ_1 (de grootste eigenwaarde) is reëel;

b) λ_1 is complex.

a) λ_1 is reëel.

Zij $\underline{x}_1$ de bijbehorende eigenkolom en $\underline{y}_1^T$ de bijbehorende eigenrij en zij bovendien $\underline{y}_1^T \underline{x}_1 = 1$.

Definieer

$$A_1 = A - \lambda_1 \underline{x}_1 \underline{y}_1^T \,. \qquad (8.11.1)$$

Bewering: als A de eigenwaarden $\lambda_1, \lambda_2, \ldots, \lambda_n$ en eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en $\underline{y}_1^T, \underline{y}_2^T, \ldots, \underline{y}_n^T$ bezit, dan heeft A_1 de eigenwaarden $0, \lambda_2, \ldots, \lambda_n$ met de eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en $\underline{y}_1^T, \underline{y}_2^T, \ldots, \underline{y}_n^T$.
Het bewijs volgt direct met gebruikmaking van $\underline{y}_1^T \underline{x}_1 = 1$, $\underline{y}_i^T \underline{x}_k = 0$ voor $k \neq i$.

Uitgaande van een willekeurige startvector $\underline{v}$ vinden we dan itererend met behulp van A_1 de eigenwaarde λ_2 en eigenvector $\underline{x}_2$.

b) λ_1 is complex.

Zij $\underline{x}_1$ en $\underline{y}_1^T$ de bijbehorende eigenvectoren en zij bovendien $\underline{y}_1^T \underline{x}_1 = 1$.
Omdat A een matrix met reële coëfficiënten is, is ook $\lambda_2 = \bar{\lambda}_1$ een eigenwaarde met bijbehorende eigenvectoren $\underline{x}_2 = \bar{\underline{x}}_1$ en $\underline{y}_2^T = \bar{\underline{y}}_1^T$.
Definieer

$$A_1 = A - \lambda_1 \underline{x}_1 \underline{y}_1^T - \lambda_2 \underline{x}_2 \underline{y}_2^T = A - 2\, R_e\, \lambda_1 \underline{x}_1 \underline{y}_1^T \,. \qquad (8.11.2)$$

Bewering: als A de eigenwaarden $\lambda_1, \lambda_2, \ldots, \lambda_n$ bezit en de eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en $\underline{y}_1^T, \underline{y}_2^T, \ldots, \underline{y}_n^T$, dan bezit A_1 de eigenwaarden $0, 0, \lambda_3, \lambda_4, \ldots, \lambda_n$ en de eigenvectoren $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en $\underline{y}_1^T, \underline{y}_2^T, \ldots, \underline{y}_n^T$.

Het bewijs gaat weer op dezelfde manier.
Uitgaande van een willekeurige vector $\underline{v}$ vinden we door iteratie met behulp van de (reële) matrix A_1 de λ_3 en $\underline{x}_3$.

Voorbeeld 7 (zie voorbeeld 5)

$$A = \begin{pmatrix} 9.08 & -1.03 & 1.04 \\ 3.99 & 4.10 & -3.92 \\ 3.97 & -3.92 & 5.94 \end{pmatrix} \qquad \lambda_1^{(1)} = 9.973448 \qquad \underline{x}_1 = \begin{pmatrix} 1.000000 \\ 0.063797 \\ 0.922268 \end{pmatrix}$$

We bepalen met behulp van A^T de eigenrij $\underline{y}_1^T$ en $\lambda_1^{(2)}$.
We vinden $\lambda_1^{(2)} = 9.973443$ $\quad \underline{y}_1^T = (0.820405, -0.811281, 1.000000)$.

Neem $\lambda_1 = \dfrac{\lambda_1^{(1)} + \lambda_1^{(2)}}{2} = 9.973446$

$\underline{y}_1^T \underline{x}_1 = 1.690916$

$A_1 = A - \lambda_1 \dfrac{\underline{x}_1 \underline{y}_1^T}{\underline{y}_1^T \underline{x}_1} = A - 5.898250\, \underline{x}_1 \underline{y}_1^T .$

$$\underline{x}_1 \underline{y}_1^T = \begin{pmatrix} 0.820405 & -0.811281 & 1.000000 \\ 0.052339 & -0.051757 & 0.063797 \\ 0.756633 & -0.748219 & 0.922268 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} 4.241046 & 3.755138 & -4.858250 \\ 3.681291 & 4.405276 & -4.296291 \\ -0.492811 & 0.493183 & 0.500233 \end{pmatrix}$$

We nemen als startvector de schatting voor $\underline{x}_2$, die uitgerekend was op pag. 143.R.

$\lambda_2 \approx 8.038253$ $\qquad$ $\underline{x}_2 \approx \begin{pmatrix} 0.987836 \\ 1.000000 \\ 0.000867 \end{pmatrix}$

	x_1	x_2	x_3
$\underline{v}$	0.987836	1.000000	0.000867
$A_1\underline{v}$	7.940384	8.038063	0.006800
$\underline{v}_1$	0.987848	1.000000	0.000846
$A_1\underline{v}_1$	7.940537	8.038197	0.006784
$\underline{v}_2$	0.987851	1.000000	0.000844
$A_1\underline{v}_2$	7.940559	8.038217	0.006781
$\underline{v}_3$	0.987851	1.000000	0.000844

Dus $\lambda_2 = 8.038217$ $\qquad$ $\underline{x}_2 = \begin{pmatrix} 0.987851 \\ 1.000000 \\ 0.000844 \end{pmatrix}$

Ingevuld in A vinden we

$$A\underline{x}_2 = \begin{pmatrix} 7.940566 \\ 8.038217 \\ 0.006782 \end{pmatrix} \qquad \lambda_2\underline{x}_2 = \begin{pmatrix} 7.940561 \\ 8.038217 \\ 0.006784 \end{pmatrix}$$

De overeenstemming is goed.

Wij kunnen dit deflatieproces voortzetten en zo successievelijk alle eigenwaarden en eigenvectoren van de matrix A bepalen.
Ook hier geldt

$$A = \lambda_1\underline{x}_1\underline{y}_1^T + \lambda_2\underline{x}_2\underline{y}_2^T + \ldots + \lambda_n\underline{x}_n\underline{y}_n^T$$

als wij zo genormeerd hebben dat $\underline{y}_i^T\underline{x}_j = \delta_{ij}$.

HOOFDSTUK IX. NUMERIEK OPLOSSEN VAN VERGELIJKINGEN

1. Inleiding

In dit hoofdstuk bespreken we eerst enkele iteratiemethoden die gebruikt kunnen worden voor het bepalen van de nulpunten van een willekeurige functie $f(x)$. Hierbij moet men reeds over een beginschatting van het nulpunt beschikken. Een dergelijke schatting kan men bijv. halen uit een grafiek. Daarna volgen enkele iteratiemethoden die geschikt zijn voor het oplossen van het stelsel vergelijkingen (niet lineair) $f(x,y) = 0$, $g(x,y) = 0$.
Tenslotte geven we nog een methode die alleen van toepassing is op polynomen.

2. Successieve substituties

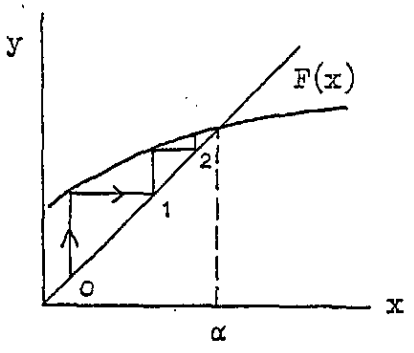
De vergelijking $f(x) = 0$ wordt geschreven in de vorm

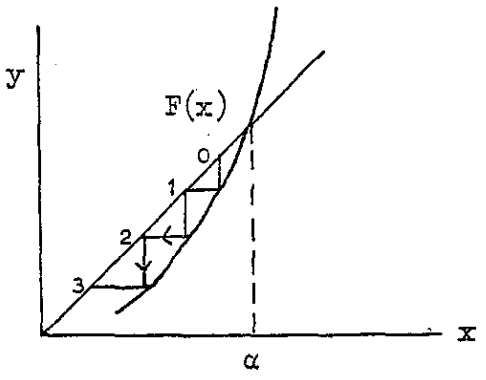
$$x = F(x). \tag{9.2.1}$$

De iteratieformule is nu

$$x_{n+1} = F(x_n). \tag{9.2.2}$$

Met (9.2.2) berekenen we, uitgaande van de beginschatting x_0, achtereenvolgens de waarden x_1, x_2, x_3, ... etc. Zij nu α de oplossing. Ligt x_0 dicht genoeg bij α en is $|F'(\alpha)| < 1$, dan convergeert de rij x_0, x_1, x_2, ... naar α.
Men kan dit met behulp van een figuur onmiddellijk inzien.





Wegens $\alpha = F(\alpha)$ volgt uit (9.2.2)

$$\alpha - x_{n+1} = F(\alpha) - F(x_n) = (\alpha - x_n)\, F'(\xi_n),$$

waarbij ξ_n tussen x_n en α ligt.

Convergeert het proces, dan is voor n voldoende groot $F'(\xi_n) \approx F'(\alpha)$ en geldt dus

$$\alpha - x_{n+1} \approx F'(\alpha)\ (\alpha - x_n), \qquad (9.2.3)$$

m.a.w. de fout in x_{n+1} is ongeveer een constante maal de fout in x_n.
Men zegt dat de convergentie lineair is of van de eerste orde. Men ziet ook weer dat voor convergentie noodzakelijk $|F'(\alpha)| < 1$ moet zijn.

Voorbeeld

Bepaal de kleinste wortel van

$$x^2 - \log x - 2 = 0; \quad \alpha \approx 0.15.$$

De vergelijking kan op verschillende manieren in de vorm $x = F(x)$ worden geschreven, bv. $x = \frac{1}{x}(\log x + 2)$, $x = \sqrt{\log x + 2}$ en $x = e^{x^2-2}$.
Alleen de laatste vorm is bruikbaar voor ons iteratieproces.
Men vindt achtereenvolgens

$$x_0 = 0.15$$
$$x_1 = 0.13842$$
$$x_2 = 0.13795$$
$$x_3 = 0.13794$$

Bij een eerste orde proces kan men uit 3 opeenvolgende iteraties een verbeterde schatting van α afleiden. Immers uit (9.2.3) volgt, voor n voldoende groot

$$\frac{\alpha - x_{n+2}}{\alpha - x_{n+1}} \approx \frac{\alpha - x_{n+1}}{\alpha - x_n}.$$

Hieruit kan men α oplossen. Dit geeft

$$\alpha \approx x_{n+2} - \frac{(x_{n+2} - x_{n+1})^2}{x_{n+2} - 2x_{n+1} + x_n} = x_{n+2} - \frac{(\nabla x_{n+2})^2}{\nabla^2 x_{n+2}} \qquad (9.2.4)$$

Men noemt deze procedure het δ^2-proces van Aitken.

Voorbeeld

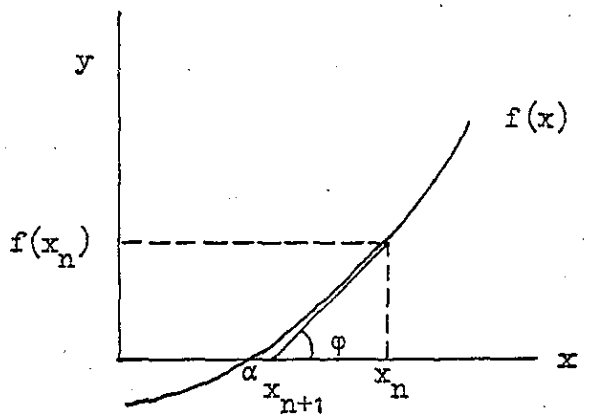
$$x^3 - 10x + 9 = 0.$$

We schrijven de vergelijking in de vorm $x = \frac{1}{10}(x^3 + 9)$ en starten met $x_0 = 1.5$.
Men vindt

$x_0 = 1.5$
$x_1 = 1.2375$
$x_2 = 1.0895$ $\quad \bar{x}_4 = 0.9989$ $\quad \bar{\bar{x}}_6 = 0.999999$
$x_3 = 1.0293$ $\quad x_5 = 0.9996704$
$x_4 = 1.0091$ $\quad x_6 = 0.9999012$

Met (9.2.4) is uit x_2, x_3 en x_4 de waarde $\bar{x}_4$ berekend; vervolgens is weer 2 maal geftereerd, waarna uit $\bar{x}_4$, x_5 en x_6 de waarde $\bar{\bar{x}}_6$ volgt.

3. Methode van Newton-Raphson



Zij x_n een schatting voor de wortel α. We trekken de raaklijn aan de kromme $y = f(x)$ in het punt $(x_n, f(x_n))$. Het snijpunt x_{n+1} van deze raaklijn met de x-as zal over het algemeen een betere benadering van de wortel zijn.
Uit de figuur lezen we af

$$f'(x_n) = \tan\varphi = \frac{f(x_n)}{x_n - x_{n+1}}.$$

Hieruit kan men x_{n+1} oplossen:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \qquad (9.3.1)$$

We gaan nog na, hoe groot de fout in x_{n+1} is. Uit (9.3.1) volgt

$$\alpha - x_{n+1} = \alpha - x_n - \frac{f(\alpha) - f(x_n)}{f'(x_n)}. \qquad (9.3.2)$$

Volgens de formule van Taylor is

$$f(\alpha) - f(x_n) = (\alpha - x_n)\, f'(x_n) + \tfrac{1}{2}(\alpha - x_n)^2\, f''(\xi_n),$$

waarbij ξ_n tussen α en x_n ligt. Substitutie in (9.3.2) geeft

$$\alpha - x_{n+1} = -\tfrac{1}{2}(\alpha - x_n)^2 \, \frac{f''(\xi_n)}{f'(x_n)}.$$

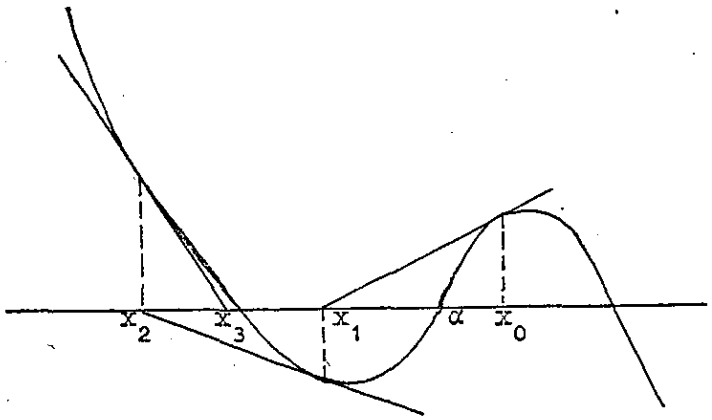
Ligt x_n dicht genoeg bij α, dan geldt dus (mits $f'(\alpha) \neq 0$)

$$\alpha - x_{n+1} \approx -\tfrac{1}{2} \, \frac{f''(\alpha)}{f'(\alpha)} \, (\alpha - x_n)^2,$$

Hieruit zien we dat, als $f'(\alpha) \neq 0$ en $f''(\alpha) \neq 0$, de fout in x_{n+1} evenredig is met het kwadraat van de fout in x_n.
Het proces convergeert kwadratisch (is van de tweede orde). Dit betekent, dat bij elke iteratie het aantal goede cijfers ongeveer wordt verdubbeld.
Als men de afgeleide van $f(x)$ gemakkelijk kan bepalen, zoals bij algebraïsche vergelijkingen, zal men wegens de snelle convergentie de voorkeur geven aan de methode van Newton-Raphson.
In andere gevallen kan men de methode van § 2 of van een van de volgende paragrafen gebruiken.
We merken nog op dat, als $f'(\alpha) = 0$, het iteratie proces van de eerste orde is.

Opmerking 1

Ook als x_0 dicht bij α ligt, behoeft het proces nog niet naar α te convergeren, zoals uit de volgende figuur blijkt.



Opmerking 2

De methoden van § 2 en § 3 zijn in principe ook bruikbaar voor complexe wortels. De moeilijkheid is hier vaak om een eerste benadering te vinden, terwijl bovendien het numeriek werken met complexe getallen onpraktisch is.

Voorbeeld

We nemen weer de vergelijking $x^3 - 10x + 9 = 0$ en starten met $x_o = 1.5$.

n	x_n	$f'(x_n)$	$f(x_n)$
0	1.5	- 3.25	- 2.63
1	0.69	- 8.57	+ 2.43
2	0.974	- 7.15399	+ 0.18401
3	0.99972	- 7.00168	+ 0.00196024
4	0.99999997		

Voorbeeld

Bepaal het reële nulpunt van $f(x) = x^k - a$, $a > 0$ $\quad$ k geheel ≥ 2.
Het iteratie proces ter bepaling van $\alpha = \sqrt[k]{a}$ wordt

$$x_{n+1} = \frac{1}{k}\left\{(k-1)x_n + \frac{a}{x_n^{k-1}}\right\} \qquad n = 0, 1, 2, \ldots\ldots$$

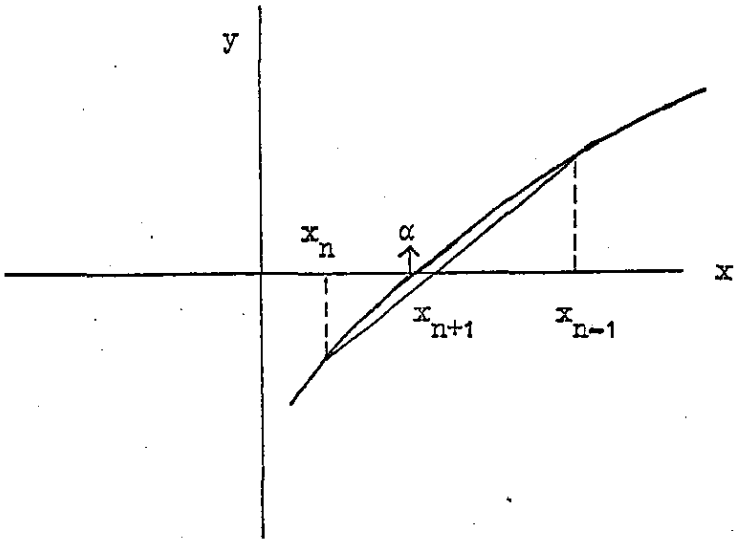
Voor k = 2 krijgt men

$$x_{n+1} = \tfrac{1}{2}\left\{x_n + \frac{a}{x_n}\right\}.$$

Dit is een gebruikelijke methode om $\sqrt{a}$ te berekenen.

4. Regula falsi

Zij x_{n-1} en x_n twee schattingen voor een wortel α van de vergelijking $f(x) = 0$. De koorde door de punten $(x_{n-1}, f(x_{n-1}))$ en $(x_n, f(x_n))$ snijdt de x-as in x_{n+1}, de nieuwe benadering voor α



In formule

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \cdot f(x_n) \qquad (9.4.1)$$

Vervolgens herhalen we dit proces met x_n en x_{n+1} enz.
Men kan bewijzen dat, als $f'(\alpha) \neq 0$ en $f''(\alpha) \neq 0$, de orde van dit proces gelijk is aan $\frac{1}{2}(1 + \sqrt{5}) = 1.618\ldots$. De convergentie is dus minder snel dan bij Newton. Maar daar staat tegenover dat men $f'(x_n)$ niet behoeft te berekenen, hetgeen bij gecompliceerde functies een wezenlijk voordeel is.

We kunnen de regula falsi ook iets anders hanteren. We bepalen eerst x_o en x_1 zodanig dat de wortel α (en geen andere) er tussen in ligt. Dit betekent dat $f(x_o) \cdot f(x_1) < 0$. Daarna berekenen wij de nieuwe benadering x_2 met de formule

$$x_2 = x_1 - \frac{x_1 - x_o}{f(x_1) - f(x_o)} \cdot f(x_1) \qquad (9.4.2)$$

Vervolgens herhalen we dit proces met x_o en x_2 of met x_1 en x_2, al naargelang $f(x_o) \cdot f(x_2) < 0$ of $f(x_1) \cdot f(x_2) < 0$ enz.
Dit proces is in het algemeen van de eerste orde. De convergentie kan worden versneld met het δ^2-proces van Aitken.

Voorbeeld

Zelfde vergelijking als in § 3. Start met $x_0 = 1.5$ en $x_1 = 0.5$. We gebruiken het eerste orde proces van de regula falsi (9.4.2)

n	x_n	$f(x_n)$
0	1.5	- 2.625
1	0.5	+ 4.125
2	1.111	- 0.739
3	1.018	- 0.12502
4	1.00276	- 0.01930
5	1.00042	

Hierbij is bv. x_4 berekend uit x_1 en x_3. Berekent men x_4 uit x_2 en x_3, dan vindt men $\bar{x}_4 = 0.99905$. Deze benadering is beter dan x_4. Men heeft dan gebruik gemaakt van formule (9.4.1).

5. Methode van Muller

Laten x_0, x_1 en x_2 drie verschillende schattingen zijn voor de wortels α van de vergelijking $f(x) = 0$. Met de interpolatie formule van Lagrange bepaalt men de parabool $P(x)$ die gaat door de drie puntenparen (x_0, f_0), (x_1, f_1) en (x_2, f_2). De wortel van de vergelijking $P(x) = 0$, die in absolute waarde het dichtst ligt bij $|x_2|$ neemt men als de nieuwe schatting x_3. Hierna berekent men f_3 en herhaalt het proces met de punten x_1, x_2 en x_3 enz. Men gaat zo lang door totdat $x_{n+1} = x_n$ in de gewenste precisie. De interpolatie formule van Lagrange voor x_0, x_1 en x_2 is

$$P(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f_2$$

Met de notatie

$$\lambda = \frac{x-x_2}{x_2-x_1} \quad , \quad a = \frac{x_2-x_1}{x_1-x_0} \quad , \quad b = a + 1$$

gaat $P(x) = 0$ over in een kwadratische vergelijking in λ.

$$\lambda^2 a\,[af_0 - bf_1 + f_2] + \lambda\,[a^2 f_0 - b^2 f_1 + (a+b)f_2] + bf_2 = 0$$

De wortels van deze vergelijking zijn

$$\lambda = \frac{-2bf_2}{c \pm \sqrt{c^2 - 4abf_2(af_0 - bf_1 + f_2)}} \qquad (9.5.1)$$

met $c = a^2 f_0 - b^2 f_1 + (a+b) f_2$

We kiezen het teken in (9.5.1) zódanig dat de absolute waarde van de breuk zo klein mogelijk is. Onze nieuwe benadering voor α volgt dan uit de betrekking: $x_3 = (x_2 - x_1)\lambda + x_2$.
De startwaarden x_0, x_1 en x_2 kunnen we bijv. halen uit een grafiek van $f(x)$.
Deze methode kan ook worden gebruikt om complexe wortels te bepaken. Het vinden van geschikte startwaarden is dan in het algemeen veel moeilijker.
De convergentie van het iteratie proces van Muller is in het algemeen van de derde orde. Bovendien behoeft men $f'(x)$ niet te berekenen hetgeen bij gecompliceerde f een wezenlijk voordeel is.
Een nadeel verbonden aan deze methode is dat men vaak met complexe arithmetiek moet rekenen ook al gaat het om het bepalen van een reëel nulpunt α van een reële functie. Het kan nl. gebeuren dat de parabool $P(x)$ geen reële snijpunten met de x-as heeft. Hierdoor is deze methode niet zo aantrekkelijk bij het gebruik van tafelrekenmachines.
We merken nog op dat, als men de wortel α in de gewenste precisie heeft bepaald, men verder kan itereren met de functie $\frac{f(x)}{(x-\alpha)}$ om de andere wortels te bepalen.

We bespreken nu enige iteratieve methoden die gebruikt kunnen worden om het stelsel (niet lineaire) vergelijkingen $f(x,y) = 0$, $g(x,y) = 0$ op te lossen.

6. Successieve substituties

Het stelsel $\left.\begin{matrix} f(x,y) = 0 \\ g(x,y) = 0 \end{matrix}\right\}$ wordt geschreven in de vorm

$$x = F(x,y), \quad y = G(x,y) \tag{9.6.1}$$

Zij (x_n, y_n) een schatting voor een wortel van (9.6.1) dan vinden we een nieuwe schatting (x_{n+1}, y_{n+1}) door te nemen

$$x_{n+1} = F(x_n, y_n), \qquad y_{n+1} = G(x_n, y_n) \tag{9.6.2}$$

We zetten dit proces voort.
Het is duidelijk dat als de rijen $\{x_n\}$ en $\{y_n\}$ limieten α, resp. β hebben (α, β) een oplossing van (9.6.1) is.
Het proces is in het algemeen van de eerste orde.

7. Methode van Newton-Raphson.

We beschouwen in R_3 de oppervlakken $z = f(x,y)$ en $z = g(x,y)$. De krommen $f(x,y) = 0$ en $g(x,y) = 0$ zijn de snijkrommen van deze oppervlakken met het x-o-y vlak. Een snijpunt (α,β) van beide krommen is een oplossing van het stelsel vergelijkingen

$$f(x,y) = 0, \quad g(x,y) = 0 \tag{9.7.1}$$

Zij (x_o,y_o) een schatting voor het punt (α,β). De raakvlakken in (x_o,y_o,f_o) en (x_o,y_o,g_o) aan de oppervlakken worden gegeven door

$$\begin{aligned} z - f_o &= \frac{\partial f}{\partial x_o}(x-x_o) + \frac{\partial f}{\partial y_o}(y-y_o) \\ z - g_o &= \frac{\partial g}{\partial x_o}(x-x_o) + \frac{\partial g}{\partial y_o}(y-y_o) \end{aligned} \tag{9.7.2}$$

Met de notatie $f_o = f(x_o,y_o)$, $g_o = g(x_o,y_o)$, $\frac{\partial f}{\partial x_o} = \left(\frac{\partial f}{\partial x}\right)_{x_o,y_o}$ enz.

De snijlijnen van beide raakvlakken met het x-o-y vlak krijgt men door in (9.7.2) z gelijk aan nul te stellen.
Men neemt het snijpunt (x_1,y_1) van deze twee snijlijnen als de nieuwe benadering van (α,β).
We vinden

$$\begin{aligned} x_1 &= x_o - \frac{f_o \frac{\partial g}{\partial y_o} - g_o \frac{\partial f}{\partial y_o}}{D} \\ &\text{met } D = \begin{vmatrix} \frac{\partial f}{\partial x_o} & \frac{\partial f}{\partial y_o} \\ \frac{\partial g}{\partial x_o} & \frac{\partial g}{\partial y_o} \end{vmatrix} \\ y_1 &= y_o - \frac{g_o \frac{\partial f}{\partial x_o} - f_o \frac{\partial g}{\partial x_o}}{D} \end{aligned} \tag{9.7.3}$$

Men zet dit proces voort totdat men (α,β) in de gewenste precisie heeft berekend. Men kan bewijzen dat dit proces in het algemeen de orde twee heeft.

8. Methode van Bairstow

We behandelen nu een methode om complexe wortels van een algebraïsche vergelijking te bepalen.
Een polynoom $P(x)$ met reële coëfficiënten is altijd te schrijven als een product van reële lineaire en kwadratische factoren. Immers zijn $a + ib$ en $a - ib$ een paar toegevoegd complexe wortels dan is de veelterm deelbaar door $(x - a - ib)(x - a + ib) = x^2 + p_o x + q_o$ met p_o en q_o reëel. Bairstow heeft een iteratieproces gegeven waarmee men een beginschatting voor p_o en q_o kan verbeteren.

Laat $x^2 + px + q$ deze schatting zijn. Wij delen $P(x)$ door $x^2 + px + q$. Dit geeft

$$P(x) = (x^2 + px + q)\, Q(x) + r_1 x + r_o \qquad (9.8.1)$$

Zij nu

$$P(x) = a_o x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$$

en

$$Q(x) = b_o x^{n-2} + b_1 x^{n-3} + \ldots + b_{n-3} x + b_{n-2}$$

dan heeft men dus

$$a_o x^n + \ldots + a_n = (x^2 + px + q)(b_o x^{n-2} + \ldots + b_{n-2}) + r_1 x + r_o .$$

Door links en rechts coëfficiënten te vergelijken vindt men

$$\begin{aligned}
a_o &= b_o \\
a_1 &= b_1 + pb_o \\
a_2 &= b_2 + pb_1 + qb_o \\
&\text{----------} \\
a_{n-2} &= b_{n-2} + pb_{n-3} + qb_{n-4} \\
a_{n-1} &= r \;\; + pb_{n-2} + qb_{n-3} \\
a_n &= r_o + qb_{n-2}
\end{aligned}$$

Hieruit kan men successievelijk b_o, b_1, ... etc. oplossen.
Definiëren we $b_{-2} = b_{-1} = 0$ dan volgt

$$b_i = a_i - pb_{i-1} - qb_{i-2} \qquad i = 0, 1, \ldots, n \qquad (9.8.2)$$

met

$$r_1 = b_{n-1},\; r_o = b_n + pb_{n-1} .$$

Relatie (9.8.2) geeft ons de coëfficiënten van het quotiëntpolynoom en de rest bij deling door $(x^2 + px + q)$.
Uit (9.8.2) volgt dat de b_i, r_o en r_1 functies zijn van p en q.
We moeten nu p en q zo trachten te bepalen dat

$$\left.\begin{aligned} r_o(p,q) &= 0 \\ r_1(p,q) &= 0 \end{aligned}\right\} \qquad (9.8.3)$$

Dit stelsel vergelijkingen lossen wij op met de iteratiemethode van Newton-Raphson.
Uitgaande van de startwaarden (p,q) vinden wij

$$\left.\begin{aligned} \tilde{p} &= p - \frac{1}{D}\left(r_1 \frac{\partial r_o}{\partial q} - r_o \frac{\partial r_1}{\partial q}\right) \\ \tilde{q} &= q - \frac{1}{D}\left(r_o \frac{\partial r_1}{\partial p} - r \frac{\partial r_o}{\partial q}\right) \end{aligned}\right. \qquad (9.8.4)$$

met

$$D = \begin{vmatrix} \frac{\partial r_1}{\partial p} & \frac{\partial r_1}{\partial q} \\ \frac{\partial r_o}{\partial p} & \frac{\partial r_o}{\partial q} \end{vmatrix}$$

$\frac{\partial r_o}{\partial p}$, $\frac{\partial r_o}{\partial q}$, $\frac{\partial r_1}{\partial p}$ en $\frac{\partial r_1}{\partial q}$ bepalen we op de volgende manier.
We gaan uit van de identiteit (9.8.1) en differentiëren links en rechts naar p en q. Dan volgt

$$\frac{\partial P}{\partial p} = 0 = x\,Q(x) + (x^2 + px + q)\frac{\partial Q}{\partial p} + \frac{\partial r_1}{\partial p}x + \frac{\partial r_o}{\partial p}$$

$$\frac{\partial P}{\partial q} = 0 = \quad Q(x) + (x^2 + px + q)\frac{\partial Q}{\partial q} + \frac{\partial r_1}{\partial q} + \frac{\partial r_o}{\partial q}$$

Hieruit volgt

$$x\,Q(x) = -\frac{\partial Q}{\partial p}(x^2 + px + q) - \frac{\partial r_1}{\partial p}x - \frac{\partial r_o}{\partial p} \qquad (9.8.5)$$

$$Q(x) = -\frac{\partial Q}{\partial q}(x^2 + px + q) - \frac{\partial r_1}{\partial q}x - \frac{\partial r_o}{\partial q} \qquad (9.8.6)$$

Uit (9.8.6) volgt dat $-\frac{\partial r_1}{\partial q}x - \frac{\partial r_o}{\partial q}$ de rest is bij deling van het polynoom Q(x) van de graad n-2 door de kwadratische factor $x^2 + px + q$.
Het quotiëntpolynoom $-\frac{\partial Q}{\partial q}$ is van de graad n-4 en is gegeven door

$$-\frac{\partial Q}{\partial q} = c_o x^{n-4} + c_1 x^{n-5} + \ldots + c_{n-4}.$$

Wij vinden dan (zie (9.8.2))

$$c_{-2} = c_{-1} = 0$$

$$c_i = b_i - pc_{i-1} - qc_{i-2} \qquad i = 0, 1, \ldots, n-2$$

met

$$-\frac{\partial r_1}{\partial q} = c_{n-3} \quad \text{en} \quad -\frac{\partial r_0}{\partial q} = c_{n-2} + pc_{n-3} \tag{9.8.7}$$

Uit (9.8.6) volgt na vermenigvuldiging met x

$$x\,Q(x) = -\left\{x\frac{\partial Q}{\partial q} + \frac{\partial r_1}{\partial q}\right\}(x^2 + px + q) + \left(p\frac{\partial r_1}{\partial q} - \frac{\partial r_0}{\partial q}\right)x + q\frac{\partial r_1}{\partial q}$$

Gelijkstelling met (9.8.5) geeft

$$p\frac{\partial r_1}{\partial q} - \frac{\partial r_0}{\partial q} = -\frac{\partial r_1}{\partial p} \quad \text{en} \quad q\frac{\partial r_1}{\partial q} = -\frac{\partial r_0}{\partial p}$$

Met (9.8.7) volgt dan

$$-\frac{\partial r_1}{\partial p} = c_{n-2} \quad \text{en} \quad -\frac{\partial r_0}{\partial p} = c_{n-1} + pc_{n-2} \tag{9.8.8}$$

met

$$c_{n-1} = -pc_{n-2} - qc_{n-3}$$

Invullen in (9.8.4) geeft

$$\tilde{p} = p + \frac{b_{n-1}c_{n-2} - b_n c_{n-3}}{D}$$

$$\tilde{q} = q + \frac{b_n c_{n-2} - b_{n-1} c_{n-1}}{D} \tag{9.8.9}$$

met

$$D = c_{n-2}^2 - c_{n-1}c_{n-3}$$

Wij vatten de procedure nog eens samen

a_o	b_o	c_o
a_1	b_1	c_1
.	.	.
.	.	.
.	.	.
.	.	.
a_{n-3}	b_{n-3}	c_{n-3}
a_{n-2}	b_{n-2}	c_{n-2}
a_{n-1}	b_{n-1}	c_{n-1}
a_n	b_n	

Bereken de kolom b's uit de a's volgens

$$b_o = a_o$$
$$b_1 = a_1 - pb_o$$
$$b_i = a_i - pb_{i-1} - qb_{i-2} \quad (i=2,..,n)$$

Bereken vervolgens de kolom c's uit de b's volgens

$$c_o = b_o$$
$$c_1 = b_1 - pc_o$$
$$c_i = b_i - qc_{i-1} - qc_{i-1} \quad (i = 2, \ldots, n-2)$$
$$c_{n-1} = - pc_{n-2} - qc_{n-3}.$$

Bereken D, $\tilde{p}$ en $\tilde{q}$ met (9.8.9).

Het proces convergeert kwadratisch als p en q dicht *genoeg* bij p_o en q_o liggen. Er treden moeilijkheden op als D nul of zeer klein wordt. Dit duidt op twee (bijna) gelijke kwadratische factoren.
Is geen beginschatting bekend, dan kan men het proces starten met bv. $p = q = 0$.
Nadat een kwadratische factor is gevonden, geven $b_o, b_1, \ldots, b_{n-2}$ de coëfficiënten van het polynoom waaruit deze factor is weggedeeld. Men kan op dezelfde manier van dit polynoom ook weer een factor bepalen, etc.
Eventueel kan men ter verhoging van de nauwkeurigheid elke kwadratische factor nog eens itereren m.b.v. de oorspronkelijke veelterm.

Voorbeeld

Los op

$$x^4 - 8x^3 + 39x^2 - 62x + 50 = (x^2 - 2x + 2)(x^2 - 6x + 25) = 0.$$

We starten met $p = q = 0$.

p,q =	0	0	-1.3	1.3	-1.9	1.9	-1.998	1.998
1	1	1	1	1	1	1		
-8	-8	-8	-6.7	-5.4	-6.10	-4.20		
39	39	39	29.0	20.7	25.51	15.63		
-62	-62	0	-15.6	33.9	-1.94	37.7		
50	50		-8.0		-2.16			
D =	1521		612		403			
$\tilde{p} - p = \Delta p =$	-1.3		-0.6		-0.098			
$q - q = \Delta q =$	1.3		0.6		0.098			

HOOFDSTUK X. APPROXIMATIE

1. Inleiding

In de numerieke wiskunde moet men vaak een functie approximeren door een andere functie, welke bestaat uit een lineaire combinatie van elementaire functies. In de berekeningen zal men dan de oorspronkelijke functie vervangen door de benaderende functie, speciaal als de waarden van deze laatste gemakkelijker te bepalen zijn en tevens binnen de vereiste nauwkeurigheid overeenkomen met de gewenste functiewaarden.

We hebben hiermee al kennis gemaakt in de eerste hoofdstukken, waar we voor deze approximatie gebruik maakten van een interpolatiepolynoom. Dit polynoom is daardoor bepaald, dat het in een aantal voorgeschreven punten dezelfde waarde heeft als de gegeven functie. Soms is het mogelijk de approximatie te verbeteren door de punten waar functie en polynoom gelijke waarde hebben niet voor te schrijven, maar op een geschikte manier aan te passen aan de gegeven functie. Dit is het geval bij de Tschebyscheff-approximatie.

Worden de functiewaarden verkregen uit metingen, waarbij dus toevallige fouten optreden, dan mag men niet te veel waarde hechten aan de afzonderlijke punten. Men zal dan een benaderende functie moeten zoeken, welke volgens een of ander criterium zo goed mogelijk past bij alle meetpunten tesamen.

Als elementaire functies kunnen behalve polynomen ook andere functies gebruikt worden. Zo ligt het voor de hand om periodieke functies te benaderen met een trigonometrische som, terwijl in andere gevallen exponentiële of rationale functies aangewezen kunnen zijn. We beperken ons tot veeltermen en trigonometrische sommen.

In dit hoofdstuk behandelen we twee methoden van approximatie, de methode van de kleinste kwadraten (zie syllabus Wiskunde II, hoofdstuk VIII) en de Tschebyscheff approximatie.

2. De methode van de kleinste kwadraten, algemeen

Gegeven is het stelsel vergelijkingen

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots a_{1k}x_k &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots a_{2k}x_k &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots a_{nk}x_k &= b_n \end{aligned} \qquad (10.2.1)$$

We veronderstellen dat $n > k$. Dan is het stelsel vergelijkingen alleen oplosbaar als er een stel van hoogstens k onderling onafhankelijke vergelijkingen is dat een oplossing heeft, terwijl de overige n-k vergelijkingen hiervan afhankelijk zijn.

In de praktijk komt het vaak voor dat men van een stel grootheden, zeg

$x_1, x_2 \ldots\ldots x_k$, een groot aantal vergelijkingen (10.2.1) kan geven, waaraan deze grootheden voldoen. Theoretisch is het stelsel vergelijkingen dan oplosbaar. Maar doordat de getallen b_i verkregen zijn uit metingen, waardoor deze getallen toevallige fouten bevatten, is het niet mogelijk het stelsel exact op te lossen. Nemen we dan precies k vergelijkingen, dan is dit stelsel wel oplosbaar, maar de oplossing kan onder invloed van de meetfouten behoorlijk ver af liggen van de werkelijke oplossing. Om de invloed van toevallige fouten in de b_i zoveel mogelijk te beperken nemen we juist een groot aantal vergelijkingen.

We veronderstellen nu dat het stelsel (10.2.1) niet oplosbaar is. Dan zal voor ieder stel waarden $x_1\ x_2 \ldots x_k$ gelden dat minstens één van de residuën

$$r_i = a_{i1}x_1 + a_{i2}x_2 + \ldots\ldots a_{ik}x_k - b_i \qquad i = 1, 2, \ldots\ldots, n \quad (10.2.2)$$

ongelijk nul is. De waarden van $x_1\ x_2 \ldots x_k$ waarvoor

$$r_1^2 + r_2^2 + \ldots\ldots r_n \text{ is minimaal} \qquad (10.2.3)$$

noemen we de oplossing van (10.2.1) volgens het principe van de kleinste kwadraten.

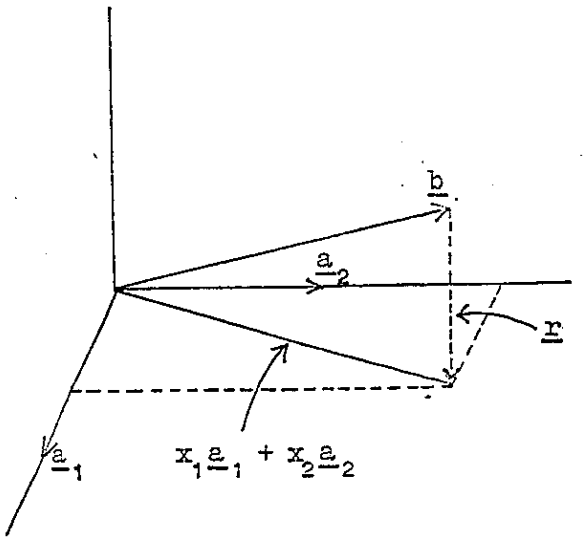
Voor een meetkundige interpretatie van dit principe nemen we een ogenblik aan dat $k = 2$ en $n = 3$.

In R_3 is de vector $\underline{r} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}$ te schrijven als

$$\underline{r} = x_1\underline{a}_1 + x_2\underline{a}_2 - \underline{b}$$

met $\underline{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}$ $\underline{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}$ $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$

Dan betekent (10.2.3) dat we zoeken naar de vector $\underline{r}$ met de kleinste lengte. Anders gezegd, we zoeken in het vlak opgespannen door $\underline{a}_1$ en $\underline{a}_2$ het punt $x_1\underline{a}_1 + x_2\underline{a}_2$ dat zo dicht mogelijk ligt bij het punt $\underline{b}$. Dit punt is de projectie van $\underline{b}$ op het vlak door $\underline{a}_1$ en $\underline{a}_2$ en de vector $\underline{r}$ staat loodrecht op dit vlak.



In het algemene geval vinden we wezenlijk hetzelfde.

In R_n noemen we

$$\underline{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} \qquad \underline{a}_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{nk} \end{pmatrix} \qquad \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \qquad \underline{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

Dan kunnen we (10.2.2) ook schrijven als

$$\underline{r} = x_1\underline{a}_1 + x_2\underline{a}_2 + \ldots x_k\underline{a}_k - \underline{b} \qquad (10.2.4)$$

De voorwaarde (10.2.3) betekent, dat we $x_1\ x_2\ \ldots\ x_k$ zo moeten bepalen dat de lengte van $\underline{r}$ minimaal is. Dit is het geval als $\underline{r}$ loodrecht staat op de lineaire deelruimte van R_n opgespannen door $\underline{a}_1\ \underline{a}_2\ \ldots\ \underline{a}_k$, dus als $\underline{r}$ loodrecht staat op ieder van de vectoren $\underline{a}_1\ \underline{a}_2\ \ldots\ \underline{a}_k$. Dit geeft voor $x_1\ x_2\ \ldots\ x_k$ de vergelijkingen

$$\begin{aligned}
(\underline{a}_1,\underline{r}) &= (\underline{a}_1,\underline{a}_1)x_1 + (\underline{a}_1,\underline{a}_2)x_2 + \ldots (\underline{a}_1,\underline{a}_k)x_k - (\underline{a}_1,\underline{b}) = 0 \\
(\underline{a}_2,\underline{r}) &= (\underline{a}_2,\underline{a}_1)x_1 + (\underline{a}_2,\underline{a}_2)x_2 + \ldots (\underline{a}_2,\underline{a}_k)x_k - (\underline{a}_2,\underline{b}) = 0 \\
&\vdots \\
(\underline{a}_k,\underline{r}) &= (\underline{a}_k,\underline{a}_1)x + (\underline{a}_k,\underline{a}_2)x_2 + \ldots (\underline{a}_k,\underline{a}_k)x_k - (\underline{a}_k,\underline{b}) = 0.
\end{aligned} \qquad (10.2.5)$$

Hier staan k vergelijkingen van de k onbekenden $x_1\ x_2\ \ldots\ x_k$, de zg. <u>normaalvergelijkingen</u>. De coëfficiëntenmatrix van dit stelsel vergelijkingen is symmetrisch; de determinant van de matrix is ongelijk aan nul als $\underline{a}_1\ \underline{a}_2\ \ldots\ \underline{a}_k$ onafhankelijk zijn. In dat geval is het stelsel oplosbaar met een van de methoden van hoofdstuk VII.

Praktisch kan men moeilijkheden krijgen als het stelsel ill-conditioned is, en dit is het geval als de vectoren $\underline{a}_1\ \underline{a}_2\ \ldots\ \underline{a}_k$ bijna afhankelijk zijn.

<u>Voorbeeld</u>

Los het volgende stelsel vergelijkingen op met de methode van de kleinste kwadraten.

$$x_1 + x_2 = 1.60$$
$$2x_1 - x_2 = 3.10$$
$$x_1 + 2x_2 = 1.62$$
$$3x_1 - 2x_2 = 4.65$$

De normaalvergelijkingen zijn

$$15x_1 - 5x_2 = 23.37$$
$$-5x_1 + 10x_2 = -7.56$$

De oplossing hiervan is

$$x_1 = 1.567 \qquad x_2 = 0.0276$$

De residuën zijn

$$r_1 = -0.0052 \quad r_2 = 0.0068 \quad r_3 = 0.0024 \quad r_4 = -0.0036$$

Opmerking 1

Als het stelsel (10.2.1) wel oplosbaar is, dan is de oplossing van (10.2.1) tevens de oplossing van (10.2.5) en de minimumlengte van de residu-vector $\underline{r}$ is nul.

Opmerking 2

Stel A is de coëfficiëntenmatrix (a_{ij}) van (10.2.1) en $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$. Dan kunnen we (10.2.1) schrijven als $A\,\underline{x} = \underline{b}$ en het stelsel normaalvergelijkingen (10.2.5) als $A^T A\,\underline{x} = A^T \underline{b}$.

Opmerking 3

De coëfficiëntenmatrix $A^T A$ van (10.2.5) hangt af van $\underline{a}_1\ \underline{a}_2 \ldots \underline{a}_k$. In het bijzonder, als deze vectoren onderling loodrecht zijn, dan wordt de matrix een diagonaalmatrix, en de oplossing van (10.2.5) vinden we dan direct, nl.

$$x_i = \frac{(\underline{b},\underline{a}_i)}{(\underline{a}_i,\underline{a}_i)} \quad \text{voor } i = 1, 2, \ldots, n.$$

3. Kleinste kwadraten met polynomen.

Veronderstel dat van de functie $f(x)$ op een segment $[a,b]$ de $(n+1)$ functie-

waarden $f(x_o)$ $f(x_1)$ ... $f(x_n)$ zijn gegeven en dat we $f(x)$ willen benaderen door een veelterm van de graad $m \leq n$,

$$y_m(x) = \lambda_o + \lambda_1 x + \ldots \lambda_m x^m,$$

volgens de methode van de kleinste kwadraten op de basis-punten $x_o, x_1, \ldots, x_n$. Dit geval is voor $m = 1$ en $m = 2$ al behandeld in de syllabus Wiskunde II. We hebben hier te maken met een bijzonder geval van § 2, waarbij namelijk $\underline{a}_i$ wordt gevormd door x^i en $\underline{b}$ door $f(x)$ op de basis punten:

$$\underline{a}_i = \begin{pmatrix} x_o^i \\ x_1^i \\ \vdots \\ x_n^i \end{pmatrix} \qquad i = 0,1,2,\ldots,m \qquad \underline{b} = \begin{pmatrix} f(x_o) \\ f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}$$

Volledigheidshalve zullen we de normaalvergelijkingen hier nogmaals afleiden op een iets andere manier. De afwijking in het punt x_j is $r_j = y_m(x_j) - f(x_j)$; de som van de kwadraten van de afwijkingen is dus

$$S = \sum_{j=0}^{n} \{y_m(x_j) - f(x_j)\}^2 = \sum_{j=0}^{n} \{\lambda_o + \lambda_1 x_j + \ldots + \lambda_m x_j^m - f(x_j)\}^2. \qquad (10.2.6)$$

De som S hangt af van de coëfficiënten $\lambda_o, \lambda_1, \ldots, \lambda_m$. Volgens het principe van de kleinste kwadraten willen we deze coëfficiënten zo kiezen dat S minimaal is.
Hiervoor is noodzakelijk dat de partiële afgeleiden $\frac{\partial S}{\partial \lambda_o}, \frac{\partial S}{\partial \lambda_1}, \ldots, \frac{\partial S}{\partial \lambda_m}$ alle nul zijn, dus

$$\frac{\partial S}{\partial \lambda_k} = -2 \sum_{j=0}^{n} \{\lambda_o + \lambda_1 x_j + \ldots + \lambda_m x_j^m - f(x_j)\} x_j^k = 0 \qquad (k=0,\ldots,m) \qquad (10.2.7)$$

Hier staan $m + 1$ lineaire vergelijkingen voor de $m + 1$ onbekende coëfficiënten.
Voeren we in de notatie

$$s_i = \sum_{j=0}^{n} x_j^i \qquad \text{en} \qquad v_i = \sum_{j=0}^{n} f(x_j)\, x_j^i ,$$

dan worden de vergelijkingen uitgeschreven

$$\begin{aligned} s_0\lambda_0 + s_1\ \lambda_1 + \ldots\ldots + s_m\ \lambda_m &= v_0 \\ s_1\lambda_0 + s_2\ \lambda_1 + \ldots\ldots + s_{m+1}\lambda_m &= v_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ s_m\lambda_0 + s_{m+1}\lambda_1 + \ldots\ldots + s_{2m}\ \lambda_m &= v_m \end{aligned} \qquad (10.2.8)$$

en dit zijn precies de normaalvergelijkingen (10.2.5) met

$$s_i = (\underline{a}_0, \underline{a}_i) = (\underline{a}_1, \underline{a}_{i-1}) = \ldots = (\underline{a}_i, \underline{a}_0)$$

en $\qquad v_i = (\underline{b}, \underline{a}_i)$

Opmerking

We hebben verondersteld $m \leqslant n$. Voor $m = n$ vinden we het interpolatie polynoom door $n + 1$ punten, dus alle afwijkingen nul.

Opmerking

De vectoren $\underline{a}_i$ zijn onderling onafhankelijk, en het stelsel normaalvergelijkingen (10.2.8) is dus altijd oplosbaar. In vele gevallen zal het stelsel echter ill-conditioned zijn. In de praktijk werkt men daarom vaak, in plaats van met x^i als elementaire functies, met orthogonale polynomen op de basispunten. Dit zijn polynomen $p_i(x)$, van de graad $i = 0, 1, 2, \ldots$, waarvoor geldt $\sum_{j=0}^{n} p_i(x_j) \cdot p_k(x_j) = 0$ als $i \neq k$.

Het benaderende polynoom wordt dan

$$y_m(x) = \mu_0 + \mu_1 p_1(x) + \ldots + \mu_m p_m(x)$$

met

$$\mu_i = \frac{\sum_{j=0}^{n} f(x_j)\, p_i(x_j)}{\sum_{j=0}^{n} p_i^{\,2}(x_j)} \qquad \text{(zie opmerking 3, blz. 168)}$$

Voorbeeld

Gegeven is onderstaande tabel

x	0.78	1.56	2.34	3.12	3.81
y	2.50	1.20	1.12	2.25	4.28

Approximeer de functie met een 2e graads veelterm

x^0	x^1	x^2	x^3	x^4	f(x)	f(x)x	$f(x)x^2$
1	0.78	0.608	0.475	0.370	2.50	1.950	1.520
1	1.56	2.434	3.796	5.922	1.20	1.872	2.921
1	2.34	5.476	12.813	29.982	1.12	2.621	6.133
1	3.12	9.734	30.371	94.759	2.25	7.020	21.902
1	3.81	14.516	55.306	210.717	4.28	16.307	62.128
5	11.61	32.768	102.761	341.750	11.35	29.770	94.604
s_0	s_1	s_2	s_3	s_4	v_0	v_1	v_2

De normaalvergelijkingen zijn dus

$$5\lambda_o + 11.61\,\lambda_1 + 32.768\lambda_2 = 11.35$$
$$11.61\,\lambda_o + 32.768\lambda_1 + 102.761\lambda_2 = 29.770$$
$$32.768\lambda_o + 102.761\lambda_1 + 341.750\lambda_2 = 94.604$$

$\lambda_o = 5.045 \qquad \lambda_1 = -4.043 \qquad \lambda_2 = 1.009.$

Dus $\quad y_2(x) = 5.045 - 4.043x + 1.009x^2.$

x	f(x)	$y_2(x)$	verschil
0.78	2.50	2.505	+0.005
1.56	1.20	1.194	-0.006
2.34	1.12	1.110	-0.010
3.12	2.25	2.252	+0.002
3.81	4.28	4.288	+0.008

4. Harmonische analyse (equidistant, discreet)

We beschouwen in deze paragraaf de approximatie van periodieke functies d.w.z. functies waarvoor geldt $f(x + p) = f(x)$ voor alle x. Men noemt p de periode. Als elementaire functies voor de approximatie gebruiken we

$\sin \frac{2k\pi x}{p}$ en $\cos \frac{2k\pi x}{p}$, $k = 0, 1, 2, \ldots$, want deze functies zijn ook periodiek met periode p.
In het volgende veronderstellen we dat $f(x)$ de periode 2 heeft. Door de substitutie $x' = \frac{2\pi}{p} x$ is dit altijd te bereiken.
Zij verder $f(x)$ gegeven in de 2n equidistante basispunten

$$x_{-n} = -\pi,\ x_{-n+1} = -\frac{(n-1)}{n}\pi,\ \ldots,\ x_{-1} = -\frac{\pi}{n},$$

$$x_0 = 0,\ x_1 = \frac{\pi}{n},\ \ldots,\ x_{n-1} = \frac{n-1}{n}\pi$$

We zullen $f(x)$ approximeren met een trigonometrische som:

$$y_m(x) = a_0 + a_1 \cos x + a_2 \cos 2x + \ldots + a_m \cos mx + b_1 \sin x + b_2 \sin 2x + \ldots + b_m \sin mx.$$

De coëfficiënten a_0 t/m a_m en b_1 t/m b_m worden volgens kleinste kwadraten zo bepaald dat

$$\sum_{j=-n}^{n-1} \{f(x_j) - y_m(x_j)\}^2 \qquad (10.4.1)$$

minimaal is.

Opmerking:

Aangezien we 2n onafhankelijke gegevens hebben, kunnen we hoogstens 2n coëfficiënten bepalen. We zullen dus eisen $m \leq n$. In het geval $m = n$ kunnen we bepalen a_0 t/m a_n en b_1 t/m b_{n-1}. De trigonometrische som heeft dan in de basispunten exact dezelfde waarden als $f(x)$, zodat we dan in feite met een trigonometrische interpolatie te doen hebben.

De berekening van de coëfficiënten wordt zeer vereenvoudigd door het feit dat de functies

$$1,\ \cos x,\ \cos 2x,\ \ldots,\ \cos (n-1)x,\ \cos nx$$
$$\sin x,\ \sin 2x,\ \ldots,\ \sin (n-1)x$$

op de basispunten een orthogonaal systeem vormen. Men kan nl. voor $0 \leq k, \ell \leq n$ de volgende relaties bewijzen

$$\sum_{j=-n}^{n-1} \sin kx_j \cos \ell x_j = 0$$

$$\sum_{j=-n}^{n-1} \cos kx_j \cos \ell x_j = \begin{cases} 0 & k \neq \ell \\ n & k = \ell \neq 0 \text{ en } \neq n \\ 2n & k = \ell = 0 \text{ of } = n \end{cases} \qquad (10.4.2)$$

$$\sum_{j=-n}^{n-1} \sin kx_j \sin \ell x_j = \begin{cases} 0 & k \neq \ell \\ n & k = \ell \neq 0 \text{ en } \neq n \\ 0 & k = \ell = 0 \text{ of } = n \end{cases}$$

We kunnen daarom de coëfficiënten direct opschrijven

$$a_o = \frac{1}{2n} \sum_{j=-n}^{n-1} f(x_j)$$

$$\left.\begin{aligned} a_k &= \frac{1}{n} \sum_{j=-n}^{n-1} f(x_j) \cos kx_j \\ b_k &= \frac{1}{n} \sum_{j=-n}^{n-1} f(x_j) \sin kx_j \end{aligned}\right\} \quad 0 < k < n \qquad (10.4.3)$$

$$a_n = \frac{1}{2n} \sum_{j=-n}^{n-1} f(x_j) \cos nx_j$$

De coëfficiënten kunnen dus door eenvoudige sommaties worden bepaald.
Men kan echter het hieraan verbonden rekenwerk nog aanzienlijk beperken.

Is de functie $f(x)$ even dan blijkt uit (10.4.3) dat alle b_k's nul zijn.
Is $f(x)$ oneven dan zijn alle a_k's nul.
Een willekeurige $f(x)$ kunnen we splitsen in een even deel en een oneven deel.
Stellen we nl.

$$F(x) = f(x) + f(-x)$$

en $$G(x) = f(x) - f(-x)$$

dan is $f(x) = \frac{1}{2} \{F(x) + G(x)\}$.

F(x) is even en G(x) is oneven.

Er volgt voor $0 < k < n$:

$$a_k = \frac{1}{n} \sum_{j=-n}^{n-1} f(x_j) \cos kx_j = \frac{1}{2n} \sum_{j=-n}^{n-1} F(x_j) \cos kx_j =$$

$$= \frac{1}{n} \left(\tfrac{1}{2}F_0 + F_1 \cos kx_1 + \ldots\ldots + F_{n-1} \cos kx_{n-1} + \tfrac{1}{2}F_n \cos kx_n\right)$$

$$b_k = \frac{1}{n} \sum_{j=-n}^{n-1} f(x_j) \sin kx_j = \frac{1}{2n} \sum_{j=-n}^{n-1} G(x_j) \sin kx_j =$$

$$= \frac{1}{n} \left(G_1 \sin kx_1 + G_2 \sin kx_2 + \ldots\ldots + G_{n-1} \sin kx_{n-1}\right),$$

en voorts

$$a_0 = \frac{1}{2n} \left(\tfrac{1}{2}F_0 + F_1 + \ldots\ldots + F_{n-1} + \tfrac{1}{2}F_n\right)$$

$$a_n = \frac{1}{2n} \left(\tfrac{1}{2}F_0 - F_1 + F_2 + \ldots\ldots + (-1)^{n-1} F_{n-1} + (-1)^n \tfrac{1}{2}F_n\right).$$

Voor n = 6 (12 basispunten) krijgen we het volgende rekenschema:

x	data	cos x	cos 2x	cos 3x	cos 4x	cos 5x	cos 6x
0	$f_0 = \frac{1}{2}F_0$	1	1	1	1	1	1
$\pi/6$	$f_1 + f_{-1} = F_1$	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{2}\sqrt{3}$	-1
$\pi/3$	$f_2 + f_{-2} = F_2$	$\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$\frac{1}{2}$	1
$\pi/2$	$f_3 + f_{-3} = F_3$	0	-1	0	1	0	-1
$2\pi/3$	$f_4 + f_{-4} = F_4$	$-\frac{1}{2}$	$-\frac{1}{2}$	1	$-\frac{1}{2}$	$-\frac{1}{2}$	1
$5\pi/6$	$f_5 + f_{-5} = F_5$	$-\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}\sqrt{3}$	-1
π	$f_6 = \frac{1}{2}F_6$	-1	1	-1	1	-1	1
	$12a_0$	$6a_1$	$6a_2$	$6a_3$	$6a_4$	$6a_5$	$12a_6$

De som van de data kolom is $12a_0$; het inproduct van de data kolom met de kolom voor cos x is $6a_1$, etc.

x	data	sin x	sin 2x	sin 3x	sin 4x	sin 5x
$\pi/6$	$f_1 - f_{-1} = G_1$	$\frac{1}{2}$	$\frac{1}{2}\sqrt{3}$	1	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$
$\pi/3$	$f_2 - f_{-2} = G_2$	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{3}$	0	$-\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{3}$
$\pi/2$	$f_3 - f_{-3} = G_3$	1	0	-1	0	1
$2\pi/3$	$f_4 - f_{-4} = G_4$	$\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{3}$	0	$\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{3}$
$5\pi/6$	$f_5 - f_{-5} = G_5$	$\frac{1}{2}$	$-\frac{1}{2}\sqrt{3}$	1	$-\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$
		$6b_1$	$6b_2$	$6b_3$	$6b_4$	$6b_5$

Voorbeeld

Gegeven is de tabel

x	$-\pi$	$-\frac{5}{6}\pi$	$-\frac{2}{3}\pi$	$-\pi/2$	$-\pi/3$	$-\pi/6$	0	$\pi/6$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π
f	1.07	1.01	1.05	1.10	1.14	1.17	1.21	1.32	1.46	1.40	1.34	1.18	1.07

Bepaal a_0 t/m a_3 en b_1 t/m b_3.

x	data	cos x	cos 2x	cos 3x	data	sin x	sin 2x	sin 3x
0	1.21	1	1	1				
$\pi/6$	2.49	0.866	0.5	0	0.15	0.5	0.866	1
$\pi/3$	2.60	0.5	-0.5	-1	0.32	0.866	0.866	0
$\pi/2$	2.50	0	-1	0	0.30	1	0	-1
$2\pi/3$	2.39	-0.5	-0.5	1	0.29	0.866	-0.866	0
$5\pi/6$	2.19	-0.866	0.5	0	0.17	0.5	-0.866	1
π	1.07	-1	1	-1				
	14.45	0.505	-0.372	-0.070		0.988	0.009	0.020

$a_0 = 1.204 \quad a_1 = 0.084 \quad a_2 = -0.062 \quad a_3 = -0.012$
$b_1 = 0.165 \quad b_2 = 0.001 \quad b_3 = 0.003$

De approximatie wordt dus

$$1.204 + 0.084 \cos x - 0.062 \cos 2x - 0.012 \cos 3x$$
$$+ 0.165 \sin x + 0.001 \sin 2x + 0.003 \sin 3x$$

Opmerking 1

Men kan het rekenschema nog verder systematiseren.
Whittaker en Robinson, The Calculus of Observations, hoofdstuk X geeft schema's voor resp. 12 en 24 basispunten.

Opmerking 2

In het voorgaande hebben we een even aantal (2n) basispunten gekozen. Men kan natuurlijk ook een oneven aantal punten nemen. Dit verandert vrijwel niets aan de theorie.

5. Harmonische analyse (continu)

Tot nu toe hebben we voor de approximatie van een functie $f(x)$ op een segment $[a,b]$ alleen gebruik gemaakt van de functiewaarden van $f(x)$ in een eindig aantal basispunten in $[a,b]$.
Het is echter ook mogelijk om de methode van de kleinste kwadraten toe te passen op alle punten van het segment. Vectoren zijn dan functies op $[a,b]$. Het ligt voor de hand om het inproduct van twee functies $u(x)$ en $v(x)$ te definiëren als volgt

$$(\underline{u}, \underline{v}) = \int_a^b u(x) \cdot (v(x)\, dx \qquad (10.5.1)$$

Ga na, dat dit inproduct voldoet aan de eigenschappen 1 t/m 6, blz. VIII. 2 van de syllabus Wiskunde II.

We zoeken in dit geval bij een functie $f(x)$ op het segment $[a,b]$ een benadering van de vorm

$$y_m(x) = \lambda_o u_o(x) + \lambda_1 u_1(x) + \ldots + \lambda_m u_m(x)$$

zodanig dat de lengte van de residu-vector, d.i. de functie

$$r(x) = f(x) - y_m(x)$$

minimaal is, dus

$$\int_a^b r^2(x)\,dx \quad \text{is minimaal.}$$

Op dezelfde wijze als in § 2 vinden we een stelsel normaalvergelijkingen

$$(\underline{u}_i, \underline{u}_o)\,\lambda_o + \ldots + (\underline{u}_i, \underline{u}_m)\,\lambda_m = (u_i, \underline{f})$$

$$i = 0, 1, 2, \ldots, m$$

waarbij de vectoren $\underline{u}_i$ de functies $u_i(x)$ en $\underline{f}$ de functie $f(x)$ op $[a,b]$ zijn. Het inproduct is nu geen som, maar de integraal (10.5.1)

We zullen als voorbeeld van een dergelijke kleinste kwadraten aanpassing behandelen, het geval van een periodieke functie $f(x)$ (periode 2π) op het segment $-\pi \leqslant x \leqslant \pi$, die we approximeren door een trigonometrische som

$$y_m(x) = a_o + \sum_{k=1}^{m} a_k \cos kx + b_k \sin kx$$

De functies cos kx en sin kx zijn orthogonaal over $(-\pi, \pi)$; er geldt

$$\int_{-\pi}^{\pi} \cos kx \sin \ell x\,dx = 0$$

$$\int_{-\pi}^{\pi} \cos kx \cos \ell x\,dx = \begin{cases} 0 & k \neq \ell \\ \pi & k = \ell \neq 0 \\ 2\pi & k = \ell = 0 \end{cases} \qquad (10.5.2)$$

$$\int_{-\pi}^{\pi} \sin kx \sin \ell x\,dx = \begin{cases} 0 & k \neq \ell \\ \pi & k = \ell \neq 0 \\ 0 & k = \ell = 0 \end{cases}$$

De residu—vector is de functie

$$r(x) = f(x) - y_m(x).$$

We moeten de lengte van deze vector minimaliseren, d.w.z. we moeten de integraal

$$I = \int_{-\pi}^{\pi} \{f(x) - y_m(x)\}^2 \, dx$$

minimaliseren. Op de bekende manier vindt men met (10.5.2) voor de coëfficiënten

$$a_o = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \, dx$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx \qquad (k > 0) \qquad (10.5.3)$$

$$b_k = \frac{1}{n} \int_{-\pi}^{\pi} f(x) \sin kx \, dx$$

Bij een gegeven functie $f(x)$ kan men ook de oneindige reeks

$$a_o + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

beschouwen, waarbij a_k en b_k gedefinieerd worden door de formules (10.5.3). Men noemt dit de Fourier-reeks.
We weten nog niet of deze reeks convergent is, en zo ja, wat dan de som is. Men kan aantonen dat voor een "nette" functie $f(x)$ de reeks inderdaad convergeert, met als som $f(x)$.
Men heeft dus

$$f(x) = a_o + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \qquad (10.5.4)$$

d.w.z. de partiële sommen $S_m(x)$ naderen tot $f(x)$ voor $m \to \infty$; volgens het voorgaande is elke partiële som tevens een approximatie van $f(x)$ in de zin der kleinste kwadraten.
De coëfficiënten a_k en b_k heten Fourier-coëfficiënten. De Fourier-reeksen spelen een belangrijke rol in de wiskunde.

Opmerking

Als men (10.5.4) links en rechts met cos kx resp. sin kx vermenigvuldigt en vervolgens integreert van $-\pi$ tot $+\pi$ (aannemende dat men de reeks term

voor term mag integreren), vindt men de uitdrukkingen (10.5.3) voor a_k en b_k weer terug.

Voorbeeld

Gegeven: $f(x) = |x|$ voor $-\pi \leqslant x \leqslant \pi$,
$f(x)$ is periodiek met periode 2π.

Ontwikkel $f(x)$ in een Fourier-reeks.

$$a_o = \frac{1}{2\pi} \int_{-\pi}^{\pi} |x| \, dx = \frac{1}{\pi} \int_{0}^{\pi} x \, dx = \tfrac{1}{2}\pi .$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \cos kx \, dx = \frac{2}{\pi} \int_{0}^{\pi} x \cos kx \, dx = \frac{2}{\pi k^2} \{(-1)^k - 1\}$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \sin kx \, dx = 0 \quad \text{(integrand is oneven functie).}$$

Men vindt aldus

$$|x| = \tfrac{1}{2}\pi - \frac{4}{\pi} \left\{\frac{\cos x}{1^2} + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \ldots\right\} \qquad \text{voor } -\pi \leqslant x \leqslant \pi.$$

Is $f(x)$ empirisch gegeven in de vorm van een grafiek of een tabel, dan moeten de Fourier-coëfficiënten langs numerieke weg worden bepaald. De praktisch meest bruikbare methode is om het probleem terug te brengen tot de in § 6 behandelde harmonische analyse voor equidistante, discrete basispunten. We verdelen het interval $(-\pi, \pi)$ in $2n$ gelijke subintervallen en berekenen met de formules (10.4.3) benaderingen voor de Fourier-coëfficiënten. Men vindt zo als benadering voor a_k

$$\bar{a}_k = \frac{1}{n} \sum_{j=-n}^{n-1} f(x_j) \cos kx_j .$$

Wegens $f(x_{-n}) = f(x_n)$ kunnen we dit nog iets anders schrijven

$$\bar{a}_k = \frac{1}{\pi} \cdot \frac{\pi}{n} \{\tfrac{1}{2}f(x_{-n}) \cos kx_{-n} + f(x_{-n+1}) \cos kx_{-n+1} + \ldots +$$

$$+ f(x_{n-1}) \cos kx_{n-1} + \tfrac{1}{2}f(x_n) \cos kx_n\}.$$

Hier staat juist de uitdrukking die men krijgt, als men $\frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx$ berekent met de trapeziumregel.

Dit lijkt een op het eerste gezicht grove benadering, die men zou kunnen verbeteren door een nauwkeurigere integratieformule, zoals Simpson, te gebruiken.
Vanwege het sterk slingerende karakter van de integrand $f(x)\cos kx$ voor grotere waarden van k, zou men het interval echter zeer fijn moeten onderverdelen. Dit levert geen praktisch bruikbare methode.

Nog een enkele opmerking over de nauwkeurigheid van de benaderingen $\bar{a}_k$ en $\bar{b}_k$. Deze hangt vooral af van de snelheid waarmee de Fourier-reeks van $f(x)$ convergeert, dus van de snelheid waarmee de echte Fourier-coëfficiënten tot nul naderen. Hierover is ons echter bij een empirisch gegeven functie niets bekend.
Verder wordt de fout groter met toenemende k (bij vaste n). In het algemeen kan men zeggen dat de besproken methode voor de lagere Fourier-coëfficiënten een redelijke benadering geeft, maar voor de hogere coëfficiënten niet.

6. Tschebyscheff-approximatie

Om de waarde van een functie $f(x)$ in een willekeurig punt x van $[a,b]$ te kunnen bepalen moeten we in vele gevallen gebruik maken van een benaderingspolynoom. Een voorbeeld hiervan is het, reeds behandelde, interpolatie polynoom $p(x)$. Nu is dit polynoom bepaald door de eis dat de waarde van het polynoom en de functie in een aantal vooraf gegeven punten van $[a,b]$ gelijk moeten zijn.

De nauwkeurigheid van de benadering wordt echter bepaald door de grootste afwijking tussen $f(x)$ en $p(x)$ op het segment $[a,b]$, d.i.

$$\max_{x \in [a,b]} |f(x) - p(x)| \qquad (10.6.1)$$

Uit het oogpunt van nauwkeurigheid is het helemaal niet interessant in welke punten $f(x)$ en $p(x)$ overeenstemmen, en men kan zich afvragen of het mogelijk is de nauwkeurigheid van de benadering te vergroten door deze punten niet vooraf voor te schrijven, maar, afhankelijk van $f(x)$, geschikt te kiezen, resp. het benaderingspolynoom op een andere wijze te bepalen waarbij de punten van overeenstemming niet expliciet gebruikt worden.

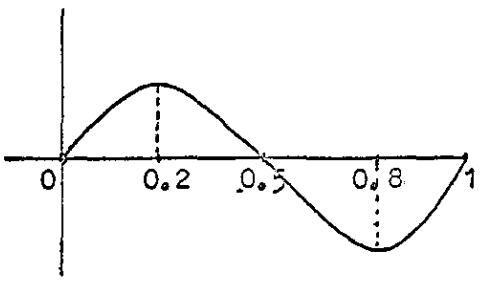
Voorbeeld 1

Het 2e graads interpolatie polynoom voor $f(x) = e^x$ op de punten 0, 0.5 en 1. van $[0,1]$ is

$$p(x) = 1 + 0.8766x + 0.8417x^2$$

maximale afwijkingen:

in x = 0.218 $\quad f(x) - p(x) =$ 0.0125
0.796 $\qquad\qquad\qquad$ -0.0147



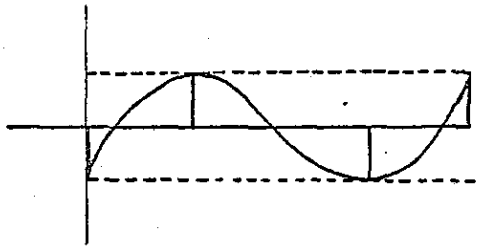
We beschouwen nu het polynoom

$$p(x) = 1.0088 + 0.8547x + 0.8461x^2$$

Dit polynoom heeft in [0.1] dezelfde waarde als f(x) in de punten 0.073, 0.520, 0.938.

Maximale afwijkingen

in x =	0.000	f(x) - p(x) = -0.0088
	0.266	0.0087
	0.766	-0.0088
	1.000	0.0088



De nauwkeurigheid van het eerste polynoom is 0.015, en van het tweede polynoom 0.009.

Als we bij het zoeken naar een benadering voor f(x) geïnteresseerd zijn in een zo groot mogelijke nauwkeurigheid bij een zo laag mogelijke graad van het polynoom, dan moeten we dus zoeken naar het polynoom p(x) waarvoor (10.6.1) zo klein mogelijk is.

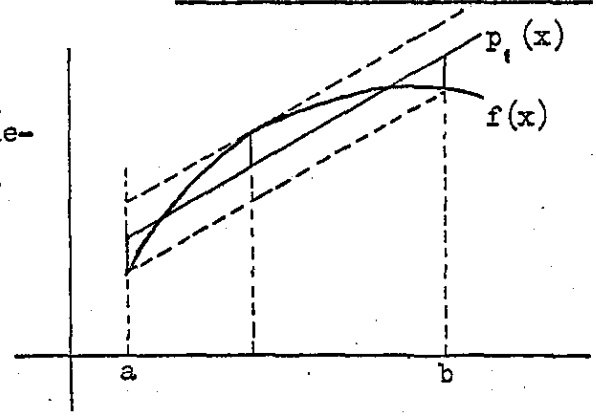
Zonder bewijs wordt hier vermeld, dat bij iedere n en iedere continue functie f(x), er precies één polynoom, zeg $p_n(x)$, bestaat waarvoor (10.6.1) minimaal is; d.w.z. voor ieder ander polynoom $g(x)$ van de graad $\leq n$ geldt

$$\max_{x \in [a,b]} |f(x) - p_n(x)| < \max_{x \in [a,b]} |f(x) - g(x)|$$

We noemen $p_n(x)$ de beste benadering van f(x) in de zin van Tschebyscheff. De approximatie waarbij als kriterium voor de beste aanpassing geldt dat de maximale afwijking (10.6.1) minimaal is, noemt men Tschebyscheff-approximatie.

Voorbeeld 2

In geval n = 1 is het beste approximatie-polynoom vaak eenvoudig te vinden, zoals uit nevenstaande tekening blijkt.



Het polynoom met de beste benadering in de zin van Tschebyscheff $p_n(x)$ is gekarakteriseerd door de volgende eigenschap, welke eveneens zonder bewijs wordt gegeven.

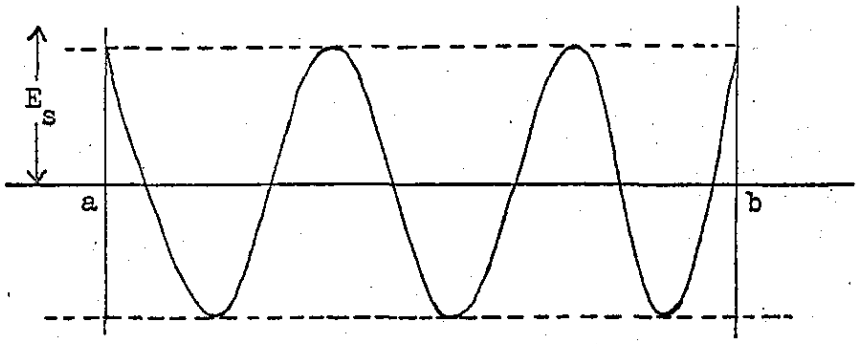
Karakteristieke eigenschap.

Er zijn in [a,b] n + 2 punten $a \leq x_0 < x, < x_2 \ldots < x_n < x_{n+1} \leq b$, waar de

fout-functie $d(x) = f(x) - p_n(x)$ de waarde E_n (d.i. het absolute maximum van $d(x)$ in $[a,b]$) aanneemt met wisselend teken, d.w.z.

$$\begin{aligned} &\text{òf} \quad d(x_i) = (-1)^i E_n \\ &\text{òf} \quad d(x_i) = (-1)^{i+1} E_n \end{aligned} \qquad i = 0, 1, 2, \ldots, n+1 \qquad (10.6.2)$$

Voorbeeld 2 demonstreert deze eigenschap voor n = 1, en voorbeeld 1 (tweede polynoom) voor n = 2.
Voor bijvoorbeeld n = 5 ziet de foutfunctie er als volgt uit (aangenomen dat er niet meer dan n + 2 extrema zijn).



De fout bij de Tschebyscheff-approximatie is dus gelijkmatig verdeeld over het gehele segment, terwijl bij interpolatie in het algemeen de fout in het midden van het segment kleiner is dan nabij de rand. (verg. de grafiek op blz. 36.R).

Het is in het algemeen zeer moeilijk om het polynoom $p_n(x)$ te vinden. Vaak zal men echter met behulp van de zogenaamde Tschebyscheff-polynomen een benadering kunnen vinden, welke niet zo ver afwijkt van de beste benadering.

Het Tschebyscheff polynoom $T_n(x)$ van de graad n wordt op het segment $-1 \leq x \leq 1$ als volgt gedefinieerd:

$$T_n(x) = \cos n\theta \qquad (10.6.3)$$

met $x = \cos \theta$.

Dat $T_n(x)$ inderdaad een veelterm van de graad n in x is, volgt uit het feit, dat we $\cos n\theta$ kunnen uitdrukken in machten van $\cos \theta$.

Volgens de formule van Euler is

$$e^{in\theta} = \cos n\theta + i \sin n\theta ,$$

dus $\cos n\theta = \mathrm{Re}\, e^{in\theta} = \mathrm{Re}(\cos \theta + i \sin \theta)^n =$

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

Vervangen we tenslotte in deze uitdrukking $\sin^2\theta$ door $1 - \cos^2\theta$, dan krijgen we een polynoom in cos θ. Zo is

$$\begin{aligned} T_0(x) &= \cos 0 = 1 \\ T_1(x) &= \cos \theta = x \\ T_2(x) &= \cos 2\theta = 2x^2 - 1 \\ T_3(x) &= \cos 3\theta = 4x^3 - 3x \\ T_4(x) &= \cos 4\theta = 8x^4 - 8x^2 + 1 \\ T_5(x) &= \cos 5\theta = 16x^5 - 20x^3 + 5x. \end{aligned} \qquad (10.6.4)$$

Uit de goniometrische formule $\cos(n+1)\theta + \cos(n-1)\theta = 2 \cos n\theta \cos \theta$ volgt onmiddellijk de recurrente betrekking

$$T_{n+1}(x) = 2x\, T_n(x) - T_{n-1}(x), \qquad (10.6.5)$$

waarmee we, uitgaande van $T_0 = 1$ en $T_1 = x$, de Tschebyscheff polynomen gemakkelijk kunnen bepalen.

Opmerking

De definitie (10.6.3) bepaalt $T_n(x)$ slechts in het segment $[-1,1]$. Als polynoom in x bestaat $T_n(x)$ natuurlijk voor alle waarden van x. Het enige echter wat we in de numerieke toepassingen van $T_n(x)$ gebruiken is zijn waarden en eigenschappen in $[-1,1]$.

Het polynoom $T_n(x)$ is in $[-1,1]$ extreem in de $n+1$ punten

$$x_k = \cos \frac{k\pi}{n} \qquad k = 0, 1, 2, \ldots, n \qquad (10.6.6)$$

en neemt in deze punten afwisselend de waarde +1 en -1 aan.

Om een functie zo goed mogelijk te benaderen, ontwikkelen we deze functie naar polynomen van Tschebyscheff. Men kan bewijzen dat iedere voldoend gladde, bv. continu differentieerbare, functie $f(x)$ in het segment $[-1,1]$ een reeks-

ontwikkeling naar polynomen van Tschebyscheff heeft, d.w.z.

$$f(x) = \sum_{k=0}^{\infty} a_k T_k(x) \qquad (10.6.7)$$

waarin

$$a_0 = \frac{1}{\pi} \int_{-1}^{1} \frac{f(x)\, T_0(x)}{\sqrt{1 - x^2}}\, dx \qquad a_k = \frac{2}{\pi} \int_{-1}^{1} \frac{f(x)\, T_k(x)}{\sqrt{1 - x^2}}\, dx \qquad (10.6.8)$$

$$k = 1, 2, 3, \ldots.$$

In feite is deze reeks niets anders dan de Fourier-reeks van de periodieke even functie $F(\theta) = f(\cos \theta)$.

We veronderstellen dat we van de functie $f(x)$ waarvan we een polynoombenadering zoeken, de reeks (10.6.7) kennen, en dat de coëfficiënten a_k snel naar nul gaan. Dit is bij veel gladde functies inderdaad het geval. Nemen we dan als benadering voor $f(x)$ de n-de partiële som van (10.6.7), d.i. het n-de graadspolynoom

$$P_n(x) = \sum_{k=0}^{n} a_k T_k(x)$$

dan is de foutfunctie $d(x) = f(x) - P_n(x)$ bij benadering gelijk aan $a_{n+1} T_{n+1}(x)$. Deze laatste functie is extreem in de $n + 2$ punten $x_k = \cos \frac{k\pi}{n+1}$, $k = 0, 1, 2, \ldots, n+1$. De waarde in deze punten is afwisselend $+ a_{n+1}$ en $- a_{n+1}$. Op grond van de karakteristieke eigenschap (blz. 181) van de beste benadering mogen we verwachten dat $P_n(x)$ bij benadering de beste approximatie is voor $f(x)$, in de zin van Tschebyscheff.

Opmerking

Voor een aantal functies bestaan er tabellen van de coëfficiënten a_n. Bijvoorbeeld, voor $f(x) = e^x$ op het segment $[-1,1]$ zijn de coëfficiënten:

$a_0 = 1.26607$
$a_1 = 1.13032$
$a_2 = .27150$
$a_3 = 4434$
$a_4 = 547$
$a_5 = 54$
$a_6 = 4$

Opmerking

De Tschebyscheff approximatie geeft geen benadering met een grote relatieve nauwkeurigheid; in de buurt van een nulpunt van $f(x)$ kan $d(x)$ juist maximaal zijn. Wil men bijvoorbeeld van $f(x) = \sin \frac{\pi}{2} x$ op $[-1,1]$ een benadering met relatieve nauwkeurigheid hebben, dan bepaalt men het benaderings polynoom niet van $f(x) = \sin \frac{\pi}{2} x$, maar van $f(x) = \dfrac{\sin \frac{\pi}{2} x}{x}$.
De benadering wordt in dit geval

$$\sin \tfrac{\pi}{2} x = x \sum a_{2k} T_{2k}(x) \quad \text{met}$$

$$\begin{aligned}
a_0 &= 2.55255\ 79248 \\
a_2 &= .28526\ 15692 \\
a_4 &= 911\ 80160 \\
a_6 &= 13\ 65875 \\
a_8 &= 11850 \\
a_{10} &= 67
\end{aligned}$$

In het geval dat we de coëfficiënten van de Tschebyscheff reeks ontwikkeling (10.6.7) niet kennen, kunnen we nog op een andere manier een zo goed mogelijke benadering trachten te vinden, nl. door het zgn. telescopen van machtreeksen. Hierbij maken we gebruik van het feit dat x^n te schrijven is als een lineaire combinatie van T_0 t/m T_n.
Met behulp van (10.6.4) vinden we

$$\begin{aligned}
1 &= T_0 \\
x &= T_1 \\
x^2 &= \tfrac{1}{2}(T_0 + T_2) \\
x^3 &= \tfrac{1}{4}(3T_1 + T_3) \\
x^4 &= \tfrac{1}{8}(3T_0 + 4T_2 + T_4) \\
x^5 &= \tfrac{1}{16}(10T_1 + 5T_3 + T_5) \\
x^6 &= \tfrac{1}{32}(10T_0 + 15T_2 + 6T_4 + T_6).
\end{aligned} \qquad (10.6.9)$$

Het procédé bij het telescopen berust hierop, dat we van een gegeven polynoom van de graad n, x^n vervangen door een lineaire combinatie van lagere machten van x, om zo de graad van het polynoom te verlagen. De fout die we hierbij maken zullen we zo klein mogelijk trachten te houden.

Voorbeeld 1.

Maak een derdegraads approximatie voor

$$f(x) = 1 - x + x^2 - x^3 + x^4 - x^5 \quad \text{op het interval } -1 \leq x \leq 1.$$

Met (10.6.9) vinden we

$$f(x) = \frac{15}{8} T_0 - \frac{19}{8} T_1 + T_2 - \frac{9}{16} T_3 + \frac{1}{8} T_4 - \frac{1}{16} T_5.$$

Dit is in feite de Tschebyscheff reeksontwikkeling (10.6.7) voor $f(x)$.
De approximatie is

$$\bar{f}(x) = \frac{15}{8} T_0 - \frac{19}{8} T_1 + T_2 - \frac{9}{16} T_3 = -\frac{9}{4} x^3 + 2x^2 - \frac{11}{16} x + \frac{7}{8}.$$

De fout is in absolute waarde kleiner dan $\frac{1}{8} + \frac{1}{16} = \frac{3}{16}$.

Men past dit procédé speciaal toe op functies waarvan de machtreeks ontwikkeling bekend is. Daarbij wordt eerst van de machtreeks een zodanige partiële som genomen dat de rest kleiner is dan de toegestane fout. Vervolgens wordt door het telescopen de graad van dit polynoom verlaagd.

Voorbeeld 2.

Approximeer e^x in $[-1,1]$ met een veelterm, zodanig dat de fout (in absolute waarde) kleiner is dan 0.01.

$$e^x = 1 + x + \tfrac{1}{2}x + \tfrac{1}{6}x^3 + \ldots.$$

Afbreken na de zesde term geeft een restterm $\frac{e^\xi}{720} x$, dus

$$|\text{fout}| = \left|\frac{e^\xi}{720} x^6\right| < \frac{e}{720} < 0.0038.$$

De veelterm $y(x) = 1 + x + \frac{1}{2} x^2 + \frac{1}{6} x^3 + \frac{1}{24} x^4 + \frac{1}{120} x^5$ voldoet dus aan de eis.

Met (10.6.9) kunnen we hiervoor schrijven

$$y(x) = \frac{81}{64} T_0 + \frac{217}{192} T_1 + \frac{13}{48} T_2 + \frac{17}{384} T_3 + \frac{1}{192} T_4 + \frac{1}{1920} T_5.$$

Laten we de laatste twee termen weg, dan geeft dit een additionele fout van $\frac{1}{192} + \frac{1}{1920} < 0.0058$.

De totale fout is dan maximaal 0.0058 + 0.0038 = 0.0096, dus nog binnen de tolerantie.
We vinden aldus

$$e^x \approx p(x) = \tfrac{81}{64} T_0 + \tfrac{217}{192} T_1 + \tfrac{13}{48} T_2 + \tfrac{17}{384} T_3 =$$

$$= 1.2656\, T_0 + 1.1302\, T_1 + 0.2708\, T_2 + 0.0443\, T_3$$

$$= 0.9948 + 0.9974\, x + 0.5417\, x^2 + 0.1771\, x^3$$

voor $-1 \leqslant x \leqslant 1$

De volgende tabel geeft een indruk van de grootte van de afwijkingen. De waarden van x zijn de punten $T_4(x)$, het eerste weggelaten Tschebyscheff-polynoom, extreem is

x	e^x	p(x)	afwijking
-1	.3679	.3629	+ 0.0059
-0.7071	.4931	.4978	- 47
0	1.0000	.9948	+ 52
0.7071	2.0281	2.0335	- 54
1	2.7183	2.7110	+ 73

HOOFDSTUK XI. SOMMATIE VAN REEKSEN

In principe kan men de som van een convergente reeks Σu_n met elke gewenste nauwkeurigheid bepalen door een voldoende aantal termen u_n uit te rekenen en bij elkaar op te tellen.
Dit proces levert geen moeilijkheden op als u_n snel genoeg naar nul gaat.
Anders wordt het echter wanneer de convergentie langzaam is.
Beschouw bijv. de reeks

$$\pi/4 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots\ldots$$

Om hiermee $\pi/4$ in bijv. 5 decimalen te berekenen, heeft men zeer veel termen nodig; de reeks is praktisch onbruikbaar. De verhouding van twee opvolgende termen nadert tot 1.

Bij alternerende reeksen is het mogelijk door een eenvoudige transformatie uit de gegeven reeks een nieuwe reeks af te leiden, die dikwijls sneller convergeert.
Zij de reeks

$$S = u_0 - u_1 + u_2 - u_3 + u_4 - u_5 + \ldots\ldots \quad (u_n > 0).$$

Hiervoor kunnen we ook schrijven

$$S = \tfrac{1}{2} u_0 - \tfrac{1}{2}(u_1 - u_0) + \tfrac{1}{2}(u_2 - u_1) - \tfrac{1}{2}(u_3 - u_2) + \ldots\ldots$$

of met de notatie $\Delta u_n = u_{n+1} - u_n$

$$S = \tfrac{1}{2} u_0 - \tfrac{1}{2}(\Delta u_0 - \Delta u_1 + \Delta u_2 - \Delta u_3 + \ldots).$$

De reeks $\Delta u_0 - \Delta u_1 + \Delta u_2 - \ldots$ kunnen we op dezelfde manier omvormen tot

$$\tfrac{1}{2} \Delta u_0 - \tfrac{1}{2}(\Delta u_1 - \Delta u_0) + \tfrac{1}{2}(\Delta u_2 - \Delta u_1) - \tfrac{1}{2}(\Delta u_3 - \Delta u_2) + \ldots$$

$$= \tfrac{1}{2} \Delta u_0 - \tfrac{1}{2}(\Delta^2 u_0 - \Delta^2 u_1 + \Delta^2 u_2 - \ldots).$$

Hiermee volgt voor S

$$S = \tfrac{1}{2} u_0 - \tfrac{1}{4} \Delta u_0 + \tfrac{1}{4}(\Delta^2 u_0 - \Delta^2 u_1 + \Delta^2 u_2 - \ldots)$$

$$= \tfrac{1}{2} u_0 - \tfrac{1}{4} \Delta u_0 + \tfrac{1}{8}\Delta^2 u_0 - \tfrac{1}{8}(\Delta^3 u_0 - \Delta^3 u_1 + \ldots) \text{ enz.}$$

Uiteindelijk vinden we zo de getransformeerde reeks

$$S = \tfrac{1}{2}u_o - \tfrac{1}{4}\Delta u_o + \tfrac{1}{8}\Delta^2 u_o - \tfrac{1}{16}\Delta^3 u_o + \tfrac{1}{32}\Delta^4 u_o - \ldots\ldots .$$

Men noemt dit proces de <u>transformatie van Euler</u>.

<u>Voorbeeld</u>

Beschouw nogmaals de reeks

$$\pi/4 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} \ldots\ldots$$

We maken een differentieschema

$$\begin{array}{lllll}
1 & & & & \\
 & -\frac{2}{3} & & & \\
\frac{1}{3} & & \frac{8}{15} & & \\
 & -\frac{2}{15} & & -\frac{16}{35} & \\
\frac{1}{5} & & \frac{8}{105} & & \frac{128}{315} \\
 & -\frac{2}{35} & & -\frac{16}{315} & \\
\frac{1}{7} & & \frac{8}{315} & & \\
 & -\frac{2}{63} & & & \\
\frac{1}{9} & & & &
\end{array}$$

De "geëulerde"reeks wordt

$$\pi/4 = \frac{1}{2} + \frac{1}{4}\cdot\frac{2}{3} + \frac{1}{8}\cdot\frac{8}{15} + \frac{1}{16}\cdot\frac{16}{35} + \frac{1}{32}\cdot\frac{128}{315} + \ldots\ldots$$

$$= \frac{1}{2} + \frac{1}{6} + \frac{1}{15} + \frac{1}{35} + \frac{4}{315} + \ldots\ldots$$

$$= \frac{1}{2}\sum_{n=0}^{\infty} \frac{n!}{1.3.5.\ \ldots(2n+1)} \cdot \ ?$$

De algemene term van de reeks volgt uit $\Delta^k u_o = \dfrac{(-1)^k\, 2^k\, k!}{1.3.\ \ldots(2k+1)}$.

De nieuwe reeks convergeert inderdaad sneller dan de oorspronkelijke. Voor grote n nadert de verhouding van twee opvolgende termen tot $\frac{1}{2}$.

Men kan nog kiezen bij welke term men met euleren zal beginnen. In het differentieschema kijkt men langs welke voorwaartse diagonaal de differenties het snelst afnemen. Tot aan die diagonaal sommeert men de termen en begint dan pas met de Euler-transformatie.

Beginnen we in ons voorbeeld na de tweede term, dan komt er

$$\pi/4 = 1 - \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{5} + \frac{1}{4} \cdot \frac{2}{35} + \frac{1}{8} \cdot \frac{8}{315} + \dots$$

$$= 1 - \frac{1}{3} + \frac{1}{10} + \frac{1}{70} + \frac{1}{315} + \dots .$$

De som van de eerste vijf termen in elk van de drie reeksen voor $\pi/4$ is resp.

$$\begin{array}{r} 0.835 \\ 0.774 \\ 0.784 \\ \pi/4 = 0.7853\dots . \end{array}$$

Opmerking

De Euler-transformatie geeft niet altijd een sneller convergente reeks. Dit zien we bijv. aan de meetkundige reeks

$$1 - r + r^2 - r^3 + \dots .$$

Het differentieschema wordt

$$\begin{array}{ccccc} 1 & & & & \\ & r-1 & & & \\ r & & (r-1)^2 & & \\ & r(r-1) & & (r-1)^3 & \\ r^2 & & r(r-1)^2 & & (r-1)^4 \\ & r^2(r-1) & & r(r-1)^3 & \\ r^3 & & r^2(r-1)^2 & & \\ & r^3(r-1) & & & \\ r^4 & & & & \end{array}$$

De geëulerde reeks wordt

$$\frac{1}{2} - \frac{1}{4}(r-1) + \frac{1}{8}(r-1)^2 - \frac{1}{16}(r-1)^3 + \dots .$$

Dit is weer een meetkundige reeks met reden $\frac{-1}{2}(r-1)$. De convergentie is verbeterd als $|\frac{1}{2}(r-1)| < |r|$, dus als $r > \frac{1}{3}$. Voor $r < \frac{1}{3}$ wordt de convergentie slechter.