Numerieke Methodes in de Algebra

Prof. Dr. Guido Vanden Berghe

Chapter 4

Numeriek berekenen van eigenwaarden en eigenvectoren van vierkante matrices

---

Doelstelling

In dit hoofdstuk wensen we de lezer te laten inzien dat het numeriek bepalen van eigenwaarden en eigenvectoren van vierkante matrices aanleiding geeft tot relatief ingewikkelde algoritmen. Drie grote klassen van methoden komen aan bod : de bepaling van eigenwaarden als wortels van de karakteristieke veelterm verbonden aan de gegeven matrix; methoden van eigenwaarde- en eigenvectorbepaling die steunen op het principe van gelijkvormigheidstransformaties; een reeks van iteratieve methoden waarmee geselecteerde eigenwaarden en geassocieerde eigenvectoren kunnen bepaald worden. Er zal op gewezen worden dat de implementatie van de besproken algoritmen zeer dikwijls met zorg en met oog voor het detail moet uitgevoerd worden. De beschikbaarheid van diagonalizatieprocedures in een stel numerieke software pakketten, voorziet in een ruim aanbod van goed geprogrammeerde algoritmen, waarvan men best zoveel mogelijk gebruik maakt.

4.1 Inleiding

De bepaling van eigenwaarden en eigenvectoren van matrices komt voor in een grote variëteit van problemen. De optredende matrices kunnen veel elementen bezitten. Ze kunnen nul zijn of ze kunnen opgebouwd zijn uit elementen die alle verschillend zijn van nul. Ze kunnen symmetrisch zijn of niet. Theoretisch kan het bepalen van eigenwaarden herleid worden tot het vinden van wortels van een algebraïsche vergelijking of tot het oplossen van een homogeen stelsel van n vergelijkingen (zie cursus Algebra). In praktische numerieke berekeningen zijn bovengenoemde methodes meestal weinig interessant en dienen andere methoden uitgewerkt te worden, die bovendien gemakkelijk tot een algoritme voor computerverwerking kunnen herleid worden.

Wanneer er een keuze mogelijk is tussen verschillende methodes, is het wel interessant vooreerst de volgende vragen te beantwoorden :

(a) zijn we geïnteresseerd in zowel eigenwaarden als eigenvectoren of zijn enkel de eigenwaarden van belang ?

(b) zijn we misschien slechts geïnteresseerd in enkele van de eigenwaarden ?

(c) bezit de matrix speciale eigenschappen ? (reëel symmetrisch, Hermitisch, etc...).

Als de eigenvectoren niet gevraagd worden zullen we uiteraard minder computer geheugenruimte nodig hebben. Als slechts enkele eigenwaarden dienen bepaald te worden, kan eventueel een speciale techniek aangewend worden. Het bepalen van alle eigenwaarden en eigenvectoren van een willekeurige niet–symmetrische matrix is zeer moeilijk, omdat deze grootheden zeer onstabiel zijn voor kleine veranderingen in de coëfficiënten van de gegeven matrix. Hierdoor zijn voor niet–symmetrische matrices zeer weinig algemene methodes en computercodes ontwikkeld. De eigenwaarden van symmetrische matrices zijn daarentegen stabiel voor storingen teweeggebracht in de matrix. Omwille hiervan en ook omwille van het feit dat dergelijke matrices frequenter optreden in realistische problemen, zijn een groot aantal methoden voor symmetrische matrices ontwikkeld. In dit hoofdstuk zullen we vooral aandacht besteden aan enkele van deze algoritmen.

Uit de cursus Algebra weten we dat λ, complex of reëel, een eigenwaarde van een vierkante matrix A is als er een vector $x \in \mathrm{C}^n,\ x \neq 0$ bestaat, zodat

$$Ax = \lambda x\ .$$

De vector x wordt dan de eigenvector corresponderend met de eigenwaarde λ genoemd. Tevens weten we dat λ een eigenwaarde is van A, als en slechts als :

$$\det(A - \lambda I_n) = 0\,. \tag{4.1}$$

Dit wordt de karakteristieke vergelijking van A genoemd. Doorgaans voert men ook nog de karakteristieke veeltermfunctie in, d.i.

$$f_A(\lambda) = \det(A - \lambda I_n)\,. \tag{4.2}$$

Is A van orde n, dan is $f_A(\lambda)$ een veelterm van de graad n, nl.

$$\begin{aligned} f_A(\lambda) \;&=\; (-1)^n\lambda^n + (-1)^{n-1}(\mathrm{tr}\ A)\lambda^{n-1} + \ldots \\ &\qquad + \text{termen van lagere graad in } \lambda\,. \end{aligned} \tag{4.3}$$

De onafhankelijke term $f_A(0)$ in die ontwikkeling is $\det A$. De overige coëfficiënten van de machten van λ in deze vergelijking zijn algebraïsch moeilijk te bepalen. Vermits $f_A(\lambda)$ van de n^{de} graad is, zijn er exact n eigenwaarden van A, als we meervoudige wortels zoveel keer mee tellen als hun multipliciteit aangeeft.

4.2 Computerbepaling van de karakteristieke veeltermfunctie Faddeev-Leverrier methode

Wanneer de orde van de matrices groter wordt is het nodig een computerkode te ontwikkelen voor de bepaling van de coëfficiënten van de karakteristieke veeltermfunctie. Eenmaal die gekend, kunnen we de wortels (= eigenwaarden van de matrix) van de vergelijking

$$f_A(\lambda) = 0$$

bepalen met één van de methoden besproken in voorgaand hoofdstuk. Verschillende methoden voor de bepaling van deze coëfficiënten zijn bekend. Hier ontwikkelen we de Faddeev–Leverrier methode die bruikbaar is voor het genereren van de veeltermfunctiecoëfficiënten behorend bij zowel *symmetrische* als bij *niet–symmetrische* matrices.

Indien de vergelijking $Ax = \lambda x$ links vermenigvuldigd wordt met A bekomen we $A^2x = \lambda Ax = \lambda^2 x$ en analoog is dan ook $A^m x = \lambda^m x$. Hieruit zien we dat als A de eigenwaarden $\lambda_1, \lambda_2, \ldots, \lambda_n$ bezit, A^m dan de eigenwaarde $\lambda_1^m, \lambda_2^m, \ldots, \lambda_n^m$ heeft. Bovendien zijn de eigenvectoren van A ook eigenvectoren van A^m. Definiëren we $S_r = \sum_{i=1}^n \lambda_i^r$ dan volgt uit het bovenstaande $\operatorname{tr} A^r = S_r$.

We weten dat de karakteristieke vergelijking van A de volgende vorm bezit :

$$f_A(\lambda) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n = 0\,.$$

Anderzijds weten we ook dat

$$f_A(\lambda) \equiv (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n)$$

zodat

$$f'_A(\lambda) = \frac{f_A(\lambda)}{\lambda - \lambda_1} + \frac{f_A(\lambda)}{\lambda - \lambda_2} + \ldots + \frac{f_A(\lambda)}{\lambda - \lambda_n}\,. \tag{4.4}$$

Het quotiënt $f_A(\lambda)/(\lambda - \lambda_i)$ wordt gemakkelijk uitgeschreven als :

$$f_A(\lambda)/(\lambda - \lambda_i) \;=\; \lambda^{n-1} + (\lambda_i + c_1)\lambda^{n-2} + (\lambda_i^2 + c_1\lambda_i + c_2)\lambda^{n-3} + \ldots \qquad (i = 1, 2, \ldots, n)\,.$$

Na sommatie van deze vergelijkingen voor $i = 1, 2, \ldots, n$, wordt (4.4) herleid tot :

$$f'_A(\lambda) = n\lambda^{n-1} + (S_1 + nc_1)\lambda^{n-2} + (S_2 + c_1S_1 + nc_2)\lambda^{n-3} + \ldots$$

Dit identificerend met $f'_A(\lambda) = n\lambda^{n-1} + (n-1)c_1\lambda^{n-2} + \ldots$, levert de $n-1$ betrekkingen :

$$\begin{aligned}
&S_1 + c_1 = 0\\
&S_2 + c_1S_1 + 2c_2 = 0\\
&\vdots\\
&S_{n-1} + c_1S_{n-2} + \ldots + c_{n-2}S_1 + (n-1)c_{n-1} = 0\,.
\end{aligned}$$

Uit $f_A(\lambda_1) + f_A(\lambda_2) + \ldots + f_A(\lambda_n) = 0$ halen we een n^{de} betrekking, nl.

$$S_n + c_1 S_{n-1} + \ldots + c_{n-1} S_1 + n c_n = 0\,.$$

Deze n betrekkingen staan in de literatuur bekend als de Newton identiteiten. De coëfficiënten van de karakteristieke vergelijking kunnen eruit bepaald worden. Vooreerst

$$c_1 = -S_1 = -\mathrm{tr}\, A$$

en dan

$$c_k = -\frac{1}{k}\Big(S_k + \sum_{j=1}^{k-1} c_j S_{k-j}\Big) \;=\; -\frac{1}{k}\Big(\mathrm{tr}\, A^k + \sum_{j=1}^{k-1} c_j \mathrm{tr}\, A^{k-j}\Big) \qquad (k = 2, 3, \ldots, n)\,. \tag{4.5}$$

Voorbeeld 4.2.1

Bepaal d.m.v. de Faddeev–Leverrier methode de eigenwaarden van de matrix

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}\,.$$

Oplossing

Uit A volgt dat

$$A^2 = \begin{bmatrix} 29 & 14 & 28 \\ 14 & 8 & 14 \\ 28 & 14 & 29 \end{bmatrix} \qquad \text{en} \qquad A^3 = \begin{bmatrix} 227 & 114 & 228 \\ 114 & 56 & 114 \\ 228 & 114 & 227 \end{bmatrix}$$

en $\mathrm{tr}\, A = 6,\ \mathrm{tr}\, A^2 = 66,\ \mathrm{tr}\, A^3 = 510$, waaruit

$$c_1 = -6$$

$$c_2 = -\tfrac{1}{2}(66 - 36) = -15$$

$$c_3 = -\tfrac{1}{3}(510 - 15 \times 6 - 6 \times 66) = -8,$$

aanleiding gevend tot de karakteristieke vergelijking

$$\lambda^3 - 6\lambda^2 - 15\lambda - 8 = 0\,.$$

De eigenwaarden zijn hieruit gemakkelijk berekenbaar vermits de veelterm gefactorizeerd kan worden, nl.

$$(\lambda - 8)(\lambda + 1)(\lambda + 1) = 0$$

zodat $\lambda_1 = 8,\ \lambda_2 = \lambda_3 = -1\,.$ ◇

Merk op dat de Faddeev–Leverrier methode bruikbaar is voor een grote klasse van matrices. De matrixelementen mogen zowel complex als reëel zijn. De te bepalen eigenwaarden mogen zowel reëel als complex zijn. Met de methode van Bairstow bv. is het mogelijk uit de karakteristieke vergelijking zowel de reële als de complexe wortels te berekenen.

DEMO 6 (Maple)

Zie Claroline

4.3 Methoden steunend op gelijkvormigheidstransformaties

Uit de cursus Algebra weten we dat vierkante matrices A en B van dezelfde orde, die met elkaar in verband staan door een *gelijkvormigheidstransformatie*, nl.

$$B = P^{-1}AP \qquad (P \text{ niet singuliere matrix}) \tag{4.6}$$

dezelfde eigenwaarden hebben. Daarenboven bestaat er een één–één–correspondentie tussen hun respectieve eigenvectoren. Als $Ax = \lambda x$ dan is $Bz = \lambda z$ met

$$z = P^{-1}x\,. \tag{4.7}$$

Er zijn een ganse reeks numerieke methoden ontwikkeld voor de bepaling van eigenwaarden en eigenvectoren steunend op bovenstaand principe. In eerste instantie werden ze ontwikkeld voor *reële symmetrische* matrices die een belangrijke rol spelen in bv. de quantummechanica. Enkele ervan zijn later uitgebreid naar Hermitische matrices, die in de moderne fysisca ook een uitermate belangrijke rol spelen. Twee van deze methoden zullen we hieronder expliciet behandelen. Reële symmetrische matrices maken deel uit van de grotere klasse Hermitische matrices. Voor dit soort matrices zijn de eigenwaarden reëel en met elke eigenwaarde kan men zoveel lineair onafhankelijke en onderling orthonormale eigenvectoren laten overeenstemmen als de multipliciteit van de eigenwaarde aangeeft.

(1) De methode van Jacobi die na een eindig aantal gelijkvormigheidstransformaties finaal resulteert in een benaderde diagonaalmatrix waar de eigenwaarden dan kunnen afgelezen worden op de hoofddiagonaal en waar de corresponderende eigenvectoren dan bijna triviaal door de methode zelf worden geleverd.

(2) De methode van Givens waarin drie grote fasen te onderscheiden zijn. De eerste bestaat uit $\frac{1}{2}(n-1)(n-2)$ orthogonale transformaties op de oorspronkelijke $n \times n$ matrix, wat aanleiding geeft tot een tridiagonale matrix. In de tweede fase wordt een rij functies gegenereerd, waarvan aangetoond wordt dat ze een Sturm rij vormen. Steunend op eigenschappen van Sturm rijen kunnen alle of een geselecteerd aantal eigenwaarden gegenereerd worden. In de derde fase worden dan de corresponderende eigenvectoren bepaald.

4.3.1 De methode van Jacobi : algemeen algoritme

In de Jacobi methode wordt de matrix P optredend in (4.6) opgebouwd als een product van een aantal speciale orthogonale matrices, m.a.w. men poogt de oorspronkelijke matrix A diagonaal te maken door een rij gelijkvormigheidstransformaties, nl. :
d.m.v. $A \to T_1^{-1}AT_1 \to T_2^{-1}(T_1^{-1}AT_1)T_2 \to T_3^{-1}(T_2^{-1}(T_1^{-1}AT_1)T_2)T_3$, enz. . .
pogen we binnen een gestelde tolerantie van afwijking een diagonale matrix D te bereiken.
De gelijkvormigheidstransformaties in de Jacobi methode worden uitgevoerd met T-matrices van het volgende type

$$T = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & & & & & & \vdots \\ & & \ddots & & & & & \\ \vdots & & & \cos\varphi & & -\sin\varphi & & 0 \\ & & & & \ddots & & & \\ 0 & & & \sin\varphi & & \cos\varphi & & 0 \\ & & & & & & \ddots & \\ 0 & & & 0 & & 0 & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ p\text{de rij} \\ \\ q\text{de rij} \\ \\ \\ \end{matrix} \qquad (4.8)$$

$$\qquad\qquad\qquad\qquad p\text{de kolom} \qquad q\text{de kolom}$$

Zulke T verschilt slechts van I_n door $t_{pp} = \cos\varphi,\ t_{pq} = -\sin\varphi,\ t_{qp} = \sin\varphi,\ t_{qq} = \cos\varphi$ met $-\pi < \varphi \leq \pi$ en $1 \leq p < q \leq n$. Voor een gegeven $n \times n$ matrix A zijn er dus $n(n-1)/2$ continue oneindige verzamelingen van mogelijke T matrices. Het is gemakkelijk na te gaan dat T^{-1} ook slechts verschillend is van I_n door $(t^{-1})_{pp} = \cos\varphi,\ (t^{-1})_{pq} = \sin\varphi,\ (t^{-1})_{qp} = -\sin\varphi,\ (t^{-1})_{qq} = \cos\varphi$.

Essentieel voor de werkwijze is de kennis van de getransformeerde matrix $B = T^{-1}AT$: (controleer die betrekkingen als oefening)

$$b_{jk} = a_{jk} \quad \text{voor} \quad \forall\, j \in [1,n] \setminus \{p,q\} \quad \text{en} \quad \forall\, k \in [1,n] \setminus \{p,q\} \qquad (4.9)$$

$$b_{jp} = b_{pj} = a_{jp}\cos\varphi + a_{jq}\sin\varphi \qquad \forall\, j \in [1,n] \setminus \{p,q\} \qquad (4.10)$$

$$b_{jq} = b_{qj} = -a_{jp}\sin\varphi + a_{jq}\cos\varphi \qquad \forall\, j \in [1,n] \setminus \{p,q\} \qquad (4.11)$$

$$b_{pp} = a_{pp}\cos^2\varphi + 2a_{pq}\sin\varphi\,\cos\varphi + a_{qq}\sin^2\varphi \qquad (4.12)$$

$$b_{qq} = a_{pp}\sin^2\varphi - 2a_{pq}\sin\varphi\,\cos\varphi + a_{qq}\cos^2\varphi \qquad (4.13)$$

$$b_{pq} = b_{qp} = -(a_{pp} - a_{qq})\sin\varphi\,\cos\varphi + a_{pq}(\cos^2\varphi - \sin^2\varphi). \qquad (4.14)$$

In de praktijk selecteert men onder de niet–diagonale elementen van A het element $a_{pq}\,, (p < q)$ waarvoor $\mid a_{pq} \mid$ maximaal is. Met deze p en q past men de hierboven beschreven Jacobi transformatie toe. De hierin optredende hoek φ wordt zo gekozen dat $b_{pq} = b_{qp} = 0$, m.a.w. uit (4.14)

$$(a_{pp} - a_{qq})\sin 2\varphi = 2a_{pq}\cos 2\varphi\ . \qquad (4.15)$$

In principe moeten we twee gevallen beschouwen om uit voorgaande betrekking φ af te leiden :

(a) $a_{pp} \neq a_{qq}$; dan is $a_{pp} - a_{qq} \neq 0$ en hieruit volgt automatisch dat ook $\cos 2\varphi \neq 0$, want ware $\cos 2\varphi = 0$ dan zou ook $(a_{pp} - a_{qq}) \sin 2\varphi = 0$, wat zou betekenen dat terzelfdertijd aan $\cos 2\varphi = 0$ en $\sin 2\varphi = 0$ zou moeten voldaan zijn, wat niet kan. Aldus volgt uit (4.15)

$$\operatorname{tg} 2\varphi = \frac{2a_{pq}}{a_{pp} - a_{qq}} = \gamma \qquad \text{met} \qquad \gamma \in \mathrm{R} \setminus \{0\} \quad \text{want} \quad a_{pq} \neq 0$$

waaruit $\varphi = \frac{1}{2}\mathrm{Bg}\operatorname{tg}\gamma$ met de restrictie $-\frac{\pi}{4} < \varphi < \frac{\pi}{4}$ om zo klein mogelijke rotaties te bewerkstelligen.
Om nu T op te bouwen zijn $\cos\varphi$ en $\sin\varphi$ vereist; deze kunnen rechtstreeks bepaald worden zonder tussenkomst van φ uit

$$\cos\varphi = \sqrt{\frac{1+\cos 2\varphi}{2}} = \sqrt{\frac{1}{2}\left(1 + \frac{1}{\sqrt{1+\operatorname{tg}^2 2\varphi}}\right)} = \sqrt{\frac{1}{2}\left(1 + \frac{1}{\sqrt{1+\gamma^2}}\right)}$$

en

$$\sin\varphi = \sqrt{\frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\gamma^2}}\right)}\operatorname{sgn}(\gamma)\,.$$

Merk op dat de tekenbepaling van $\sin\varphi$ verbonden is aan het teken van φ of het teken van γ.

(b) $a_{pp} = a_{qq}$; uit (4.15) volgt $2a_{pq}\cos 2\varphi = 0$ of $\cos 2\varphi = 0$, waaruit $\varphi = \frac{\pi}{4}$ en $\cos\varphi = \sin\varphi = \frac{\sqrt{2}}{2}$.
Het algoritme wordt nu verder opgebouwd door onder de niet–diagonale elementen van B het element $b_{p'q'}\,, (p' < q')$ te selecteren waarvoor $\mid b_{p'q'} \mid$ maximaal is. Met deze p' , q' past men een nieuwe Jacobi transformatie toe met zulke hoek φ dat in de matrix $C = T_2^{-1}BT_2$, $c_{p'q'} = c_{q'p'} = 0$ wordt, enz...

Indien geen interferentie zou plaats vinden tussen de opeenvolgende respectieve Jacobi transformaties, zou A in ten hoogstens $n(n-1)/2$ (= aantal niet–diagonale elementen boven de diagonaal in A) Jacobi transformaties diagonaal zijn. Evenwel worden door dergelijke transformaties met bepaalde p en q alle elementen van de p^{de} en q^{de} rij (kolom) van de getransformeerde matrix gewijzigd (zie (4.10), (4.11)). Vandaar dat een element a_{pq}, $(p < q)$ dat tot nul herleid werd, eventueel weer verschillend van nul kan worden door de perturberende invloed van een andere Jacobi transformatie. Spijts deze schijnbaar vicieuze cirkel is de methode van Jacobi finaal toch convergent.

4.3.2 Convergentie bij de methode van Jacobi

Vooreerst willen we opmerken dat uit de omzettingsformules (4.9)–(4.14) een reeks invarianten van de transformatie kunnen afgeleid worden

(a) Uit de optelling van (4.12) en (4.13) :

$$b_{pp} + b_{qq} = a_{pp} + a_{qq} \tag{4.16}$$

(b) Uit de aftrekking van (4.12) en (4.13) en uit (4.14) :

$$b_{pp} - b_{qq} = (a_{pp} - a_{qq}) \cos 2\varphi + 2a_{pq} \sin 2\varphi$$
$$2b_{pq} = -(a_{pp} - a_{qq}) \sin 2\varphi + 2a_{pq} \cos 2\varphi$$

De som der kwadraten van beide betrekkingen levert :

$$(b_{pp} - b_{qq})^2 + 4b_{pq}^2 = (a_{pp} - a_{qq})^2 + 4a_{pq}^2 \tag{4.17}$$

(c) Uit de som van (4.17) en het kwadraat van (4.16) mits deling door twee :

$$b_{pp}^2 + b_{qq}^2 + 2b_{pq}^2 = a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 \tag{4.18}$$

(d) Tenslotte volgt uit (4.9)–(4.11) en (4.18) dat :

$$\sum_{j=1}^{n} \sum_{k=1}^{n} b_{jk}^2 = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}^2 \,. \tag{4.19}$$

Bij toepassing van de eerste Jacobi transformatie die $a_{pq} = a_{qp} \neq 0$ omzet naar $b_{pq} = b_{qp} = 0$ volgt uit (4.18) :

$$b_{pp}^2 + b_{qq}^2 = a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2$$

en uit (4.9)

$$b_{jj}^2 = a_{jj}^2 \quad \text{voor } \forall\, j \in [1, n] \setminus \{p, q\} \,.$$

Vandaar ook

$$\sum_{j=1}^{n} b_{jj}^2 = \sum_{j=1}^{n} a_{jj}^2 + 2a_{pq}^2 \,. \tag{4.20}$$

Door nu van (4.20) van (4.19) af te trekken bekomen we

$$\sum_{\substack{j=1 \\ (j \neq k)}}^{n} \sum_{k=1}^{n} b_{jk}^2 = \sum_{\substack{j=1 \\ (j \neq k)}}^{n} \sum_{k=1}^{n} a_{jk}^2 - 2a_{pq}^2$$

of ook wegens het symmetrisch zijn van A en B

$$\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} b_{jk}^2 = \sum_{j=1}^{n-1}\sum_{k=j+1}^{n} a_{jk}^2 - a_{pq}^2 \,, \tag{4.21}$$

d.w.z. de som van de kwadraten van de elementen boven de hoofddiagonaal in de getransformeerde matrix B is gelijk aan de analoge som in A verminderd met a_{pq}^2. Noteren we nu die som in A als s_0 en geeft s_m de som van de kwadraten van deze zelfde elementen aan in de $n \times n$ matrix, die ontstaan is na m Jacobi transformaties van het geschetste type, dan is (4.21) herschrijfbaar als

$$s_1 = s_0 - a_{pq}^2 \,. \tag{4.22}$$

Wegens $\mid a_{pq} \mid$ maximaal is het evident dat

$$a_{pq}^2 \geq \frac{s_0}{\frac{n(n-1)}{2}} = \frac{s_0}{N} \,, \tag{4.23}$$

wat wil zeggen dat het kwadraat van a_{pq} niet kleiner kan zijn dan de gemiddelde waarde van het kwadraat der elementen boven de hoofddiagonaal. Vandaar dat uit (4.22) en(4.23) volgt

$$0 \leq s_1 \leq s_0 - \frac{s_0}{N} \qquad \text{of} \qquad 0 \leq s_1 \leq s_0(1 - \frac{1}{N}) \,.$$

Analoge ongelijkheden behoren bij elke volgende transformatie, d.w.z.

$$\begin{aligned} 0 \leq s_2 \leq (1 - \frac{1}{N})s_1 &\leq (1 - \frac{1}{N})^2 \, s_0 \\ 0 \leq s_3 \leq (1 - \frac{1}{N})s_2 &\leq (1 - \frac{1}{N})^3 \, s_0 \end{aligned}$$

of na m transformaties :

$$0 \leq s_m \leq (1 - \frac{1}{N})^m \, s_0 \qquad \text{of} \qquad 0 \leq s_m \leq s_0 e^{m \, \ln(1-1/N)}$$

wat wegens

$$\ln(1 - \frac{1}{N}) = -\frac{1}{N} - \frac{1}{2N^2} - \ldots \qquad\qquad (0 < \frac{1}{N} < 1)$$

te schrijven is als $0 \leq s_m < s_0 \, e^{-m/N}$.
Dus de som van de kwadraten van de niet–diagonaalelementen daalt op exponentiële wijze bij toenemende m. Dus

$$\lim_{m\to\infty} s_m = \lim_{m\to\infty} \sum_{j=1}^{n-1}\sum_{k=j+1}^{n} (a_{jk}^{(m)})^2 = 0$$

waarbij $a_{pq}^{(m)}$ het (p,q)–element van A representeert na m transformaties. Hieruit volgt uiteraard

$$\lim_{m\to\infty} a_{jk}^{(m)} = 0 \qquad\qquad (j \neq k)\ . \tag{4.24}$$

Dus elk niet–diagonaalelement convergeert naar nul na een relatief groot aantal Jacobi transformaties.

Om nu de convergentie van elk diagonaalelement naar een limiet (= een eigenwaarde) te onderzoeken, beschouwen we de matrix $[a_{jk}^{(m)}]$ bekomen na m transformaties. Zij daarenboven p en q de rangnummers van rij en kolom die bij de $(m+1)^{\text{de}}$ transformatie veranderd worden, dan volgt uit (4.9), (4.12) en (4.13) :

$$\begin{aligned}
a_{jj}^{(m+1)} &= a_{jj}^{(m)} \qquad\qquad \forall\, j \in [1,n] \setminus \{p,q\} \\
a_{pp}^{(m+1)} &= a_{pp}^{(m)} \cos^2\varphi + 2a_{pq}^{(m)} \cos\varphi \,\sin\varphi + a_{qq}^{(m)} \sin^2\varphi \qquad (4.25)\\
a_{qq}^{(m+1)} &= a_{pp}^{(m)} \sin^2\varphi - 2a_{pq}^{(m)} \cos\varphi \,\sin\varphi + a_{qq}^{(m)} \cos^2\varphi\ . \qquad (4.26)
\end{aligned}$$

Beschouwen we in het bijzonder het p^{de} diagonaalelement. Uit (4.25) deduceren we

$$a_{pp}^{(m+1)} - a_{pp}^{(m)} = -(a_{pp}^{(m)} - a_{qq}^{(m)}) \sin^2\varphi + 2a_{pq}^{(m)} \cos\varphi \,\sin\varphi\ .$$

Uit voorgaande paragraaf volgt :

(a) als $a_{pp}^{(m)} \neq a_{qq}^{(m)}$, dat

$$\operatorname{tg} 2\varphi = \frac{2a_{pq}^{(m)}}{a_{pp}^{(m)} - a_{qq}^{(m)}} \quad \left(-\frac{\pi}{4} < \varphi < \frac{\pi}{4}\right)$$

wat, gesubstitueerd in voorgaande uitdrukking, leidt tot :

$$\begin{aligned}
a_{pp}^{(m+1)} - a_{pp}^{(m)} &= -2a_{pq}^{(m)} \operatorname{cotg} 2\varphi \,\sin^2\varphi + 2a_{pq}^{(m)} \cos\varphi \,\sin\varphi \\
&= 2a_{pq}^{(m)} \sin\varphi \left[-\frac{1-\operatorname{tg}^2\varphi}{2\operatorname{tg}\varphi}\,\sin\varphi + \cos\varphi\right] \\
&= 2a_{pq}^{(m)} \sin\varphi\,\cos\varphi \frac{1+\operatorname{tg}^2\varphi}{2} = a_{pq}^{(m)} \operatorname{tg}\varphi
\end{aligned}$$

(b) als $a_{pp}^{(m)} = a_{qq}^{(m)}$,dat $\varphi = \dfrac{\pi}{4}$
waaruit volgt

$$a_{pp}^{(m+1)} - a_{pp}^{(m)} = a_{pq}^{(m)}\ .$$

Zowel uit (a) en (b) kunnen we afleiden dat

$$\mid a_{pp}^{(m+1)} - a_{pp}^{(m)} \mid \leq \mid a_{pq}^{(m)} \mid \ . \tag{4.27a}$$

Op analoge wijze vertrekkend uit (4.26) kan men bewijzen dat

$$\mid a_{qq}^{(m+1)} - a_{qq}^{(m)} \mid \leq \mid a_{pq}^{(m)} \mid \ . \tag{4.27b}$$

Omdat alle niet–diagonaalelementen voor grote m (veel Jacobi transformaties) exponentieel tot nul naderen (zie (4.24)) volgt uit (4.27) dat elk diagonaalelement een unieke limiet bezit.

In de praktijk dient het rekenproces binnen een eindige tijdsspanne gestopt te worden. Gezien echter iedere Jacobi transformatie dient voorafgegaan te worden door het zoeken van het niet–diagonaalelement a_{pq} met maximale absolute waarde, kan, éénmaal dit element gevonden, $\mid a_{pq} \mid$ vergeleken worden met een vooropgegeven tolerantie ϵ t.o.v. nul. Indien deze $\mid a_{pq} \mid < \epsilon$ dienen geen verdere Jacobi transformaties meer verricht te worden, en zijn de diagonaalelementen de te zoeken eigenwaarden. Indien $\mid a_{pq} \mid > \epsilon$ zetten we het algoritme verder.

4.3.3 Bepaling van de eigenvectoren in de Jacobi methode

De methode van Jacobi is van zulkdanige efficiëntie dat ze toelaat, praktisch zonder supplementaire berekeningen, een stel op één genormeerde orthogonale reële eigenvectoren van A te bekomen, één bij elke benaderde eigenwaarden. Zij $T_1,\ T_2,\ \ldots,\ T_a$ de opeenvolgende Jacobi transformatiematrices waarmee A behandeld werd, om benaderend omgezet te worden tot een diagonaalmatrix D, nl.

$$T_a^{-1}\, T_{a-1}^{-1} \ldots T_2^{-1}\, T_1^{-1} A T_1 T_2 \ldots T_{a-1} T_a \cong D$$

met D een exacte diagonaalmatrix met de eigenwaarden van A in een bepaalde volgorde op de hoofddiagonaal. Voor de matrix D kan men het volgende stel op één genormeerde orthogonale eigenvectoren kiezen :

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad , \qquad x_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad , \qquad \ldots \quad , \qquad x_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \ .$$

Ze voldoen aan $x_j^\dagger x_k = \delta_{jk}$. Uit $D x_k = \lambda_k x_k$ volgt bij benadering

$$T_a^{-1}\, T_{a-1}^{-1} \ldots T_2^{-1}\, T_1^{-1}\, A T_1 T_2 \ldots T_{a-1} T_a x_k \cong \lambda_k x_k \ .$$

Linkse vermenigvuldiging van beide leden met achtereenvolgens T_a , T_{a-1} , $\ldots$, T_1 geeft wegens $T_s T_s^{-1} = 1$,

$$A T_1 T_2 \ldots T_{a-1} T_a x_k \cong \lambda_k T_1 T_2 \ldots T_{a-1} T_a x_k \, .$$

Hieruit volgt dat $T_1T_2\ldots T_{a-1}T_ax_k$ een benaderde eigenvector is van A behorend bij λ_k, d.w.z. elke kolom in $T_1T_2\ldots T_{a-1}T_a$ representeert een welbepaalde eigenvector. Deze benaderde eigenvectoren zijn op één genormeerd en onderling orthogonaal, want

$$\begin{aligned}&(T_1T_2\ldots T_ax_j)^\dagger(T_1T_2\ldots T_ax_k)\\&\quad=x_j^\dagger T_a^\dagger\ldots T_2^\dagger T_1^\dagger T_1T_2\ldots T_ax_k\\&\quad=x_j^\dagger T_a^{-1}\ldots T_2^{-1}T_1^{-1}T_1T_2\ldots T_ax_k=x_j^\dagger x_k=\delta_{jk}\,.\end{aligned}$$

4.3.4 De methode van Givens : algemeen algoritme

Net zoals de Jacobi methode is de Givens methode geschikt voor reële symmetrische matrices. Opnieuw wordt door een opeenvolging van gelijkvormigheidstransformaties de oorspronkelijke matrix omgevormd, nu echter niet tot een diagonale matrix, maar tot een tridiagonale matrix. Hiertoe maakt men eveneens gebruik van de orthogonale matrices van het type (4.8) . Nochtans wordt hier een specifieke volgorde in de uit te voeren transformaties opgelegd. We beschouwen vooreerst een transformatie T_1 van het type (4.8) met $p=2$ en $q=3$. Zulke transformatie wordt ook een rotatie in het (2,3)-vlak genoemd. De gelijkvormigheidstransformatie $A\to A^{(1)}=T_1^{-1}AT_1$ beïnvloedt alle elementen in de tweede en de derde rij en in de tweede en in de derde kolom. De optredende hoek φ wordt nu bepaald door de voorwaarde $a_{13}^{(1)}=a_{31}^{(1)}=0$, en niet zoals in de Jacobi methode door de voorwaarde $a_{23}^{(1)}=a_{32}^{(1)}=0$. Uit (4.11) volgt dat met de huidige notatie deze voorwaarde te schrijven is als

$$a_{13}^{(1)}=-a_{12}\sin\varphi+a_{13}\cos\varphi=0\,,$$

waaruit

$$\mathrm{tg}\,\varphi=a_{13}/a_{12}\quad,(-\frac{\pi}{2}<\varphi<\frac{\pi}{2})\quad\text{als } a_{12}\neq 0$$

en

$$\varphi=\frac{\pi}{2}\quad\text{als } a_{12}=0.$$

De volgende gelijkvormigheidstransformatie $A^{(1)}\to A^{(2)}=T_2^{-1}A^{(1)}T_2$ wordt uitgevoerd met een rotatie T_2 in het (2,4)-vlak over een hoek φ die wordt bepaald door de voorwaarde $a_{14}^{(2)}=a_{41}^{(2)}=0$, wat resulteert in

$$\mathrm{tg}\,\varphi=a_{14}^{(1)}/a_{12}^{(1)}=a_{14}/a_{12}^{(1)}\quad\text{als } a_{12}^{(1)}\neq 0$$

en

$$\varphi=\frac{\pi}{2}\quad\text{als } a_{12}^{(1)}=0.$$

Bij deze tweede gelijkvormigheidstransformatie veranderen enkel de elementen van de tweede en vierde rij, respectievelijk kolom, zodat in het bijzonder $a_{13}^{(2)} = a_{13}^{(1)} = 0$. Op analoge wijze worden nu achtereenvolgens m.b.v. rotatie $T_3, \ldots, T_{n-2}$ respectievelijk in het (2,5)-vlak, ..., (2,n)-vlak, alle volgende elementen op de eerste rij nul gemaakt, nl.

$$a_{13}^{(n-2)} = a_{14}^{(n-2)} = \ldots = a_{1n}^{(n-2)} = 0.$$

Vervolgens wordt gebruik gemaakt van rotaties in de vlakken (3,4), (3,5), ..., (3,n) om telkens een element uit de tweede rij nul te maken. We moeten onderzoeken wat daarbij gebeurt met de elementen uit de eerste rij die reeds nul gesteld werden. Beschouwen we bijvoorbeeld de rotatie T_{n-1} in het (3,4)-vlak over een hoek φ bepaald door de voorwaarde $a_{24}^{(n-1)} = 0$. De gelijkvormigheidstransformatie beïnvloedt alle elementen in de derde en vierde rij, respectievelijk kolom. Door te steunen op (4.10) en (4.11) leiden we af dat

$$\begin{aligned} a_{13}^{(n-1)} &= a_{13}^{(n-2)} \cos\varphi + a_{14}^{(n-2)} \sin\varphi = 0 \\ a_{14}^{(n-1)} &= -a_{13}^{(n-2)} \sin\varphi + a_{14}^{(n-2)} \cos\varphi = 0\,, \end{aligned}$$

m.a.w. de nulelementen op de eerste rij blijven onveranderd.

Uit het voorgaande is eenvoudig te begrijpen hoe de procedure kan verdergezet worden om finaal een bandmatrix B te bekomen (d.w.z. $b_{ij} = 0$ als $|i-j| > 1$), nl. :

$$B = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 & \ldots & 0 \\ \beta_1 & \alpha_2 & \beta_2 & 0 & \ldots & 0 \\ 0 & \beta_2 & \alpha_3 & \beta_3 & \ldots & 0 \\ & & & \ddots & & \vdots \\ & & & & \ddots & \vdots \\ 0 & 0 & \ldots & \ldots & \beta_{n-1} & \alpha_n \end{bmatrix}, \tag{4.28}$$

met B resulterend uit de transformatie

$$B = (T_1 T_2 \ldots T_s)^{-1} A (T_1 T_2 \ldots T_s)$$

waarbij $s = \frac{1}{2}(n-1)(n-2)$.

Vervolgens zullen we nu de karakteristieke veeltermfunctie van B bepalen en deze verder gebruiken voor de berekening van de eigenwaarden van B. Om

$$f_n(\lambda) = \det(B - \lambda I)$$

af te leiden, definiëren we de rij $\{f_k(\lambda) | 0 \leq k \leq n\}$ met

$$f_k(\lambda) = \det \begin{bmatrix} \alpha_1 - \lambda & \beta_1 & 0 & \ldots & 0 \\ \beta_1 & \alpha_2 - \lambda & \beta_2 & \ldots & \vdots \\ 0 & & \ddots & & \vdots \\ \vdots & & & \ddots & \beta_{k-1} \\ 0 & \ldots & \ldots & \beta_{k-1} & \alpha_k - \lambda \end{bmatrix}, \tag{4.29}$$

en $f_0(\lambda) = 1$. Uit rechtstreekse berekening volgt :

$$\begin{aligned} f_1(\lambda) &= \alpha_1 - \lambda \\ f_2(\lambda) &= (\alpha_2 - \lambda)(\alpha_1 - \lambda) - \beta_1^2 \\ &= (\alpha_2 - \lambda) f_1(\lambda) - \beta_1^2 f_0(\lambda) \,. \end{aligned}$$

De formule voor $f_2(\lambda)$ illustreert de algemene recursierelatie waaraan de elementen van de rij voldoen, nl.

$$f_k(\lambda) = (\alpha_k - \lambda) f_{k-1}(\lambda) - \beta_{k-1}^2 \, f_{k-2}(\lambda) \;, \qquad 2 \leq k \leq n \;. \tag{4.30}$$

Het resultaat (4.30) volgt rechtstreeks uit de ontwikkeling van de determinant (4.29) naar zijn laatste rij m.b.v. minoren. Op dit punt kunnen we het eigenwaardenprobleem als opgelost beschouwen, vermits $f_n(\lambda)$ op een gemakkelijke wijze uit (4.30) recursief kan opgebouwd worden. De te zoeken eigenwaarden zijn dan de wortels van

$$f_n(\lambda) = 0 \,,$$

en deze kunnen met één van de methoden besproken in hoofdstuk 3 bepaald worden. Nochtans bezit de rij $\{f_k(\lambda) \mid 0 \leq k \leq n\}$ bijzondere eigenschappen, waardoor het een zgn. Sturm rij wordt; door gebruik te maken van deze eigenschappen wordt het relatief eenvoudig om bepaalde eigenwaarden van B te isoleren en dan enkel deze te bepalen m.b.v. een iteratieve methode. De speciale eigenschappen die van $\{f_k(\lambda)\}$ een Sturm rij maken zullen in volgende paragraaf toegelicht worden.

Nochtans willen we hier vooreerst nog de situatie beschouwen waarbij in B één van de β_l gelijk nul wordt, bijvoorbeeld we verkrijgen een tridiagonale matrix van de vorm

$$B = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 \\ \beta_1 & \alpha_2 & 0 & 0 & 0 \\ 0 & 0 & \alpha_3 & \beta_3 & 0 \\ 0 & 0 & \beta_3 & \alpha_4 & \beta_4 \\ 0 & 0 & 0 & \beta_4 & \alpha_5 \end{bmatrix} .$$

Definiëren we nu B_1 en B_2 als de twee blokken langs de diagonaal van orde 2 en 3 respectievelijk, m.a.w.

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} .$$

Hieruit volgt dat

$$\det[B - \lambda I_5] = \det[B_1 - \lambda I_2] \det[B_2 - \lambda I_3]$$

en we kunnen de eigenwaarden van B vinden door deze van B_1 en B_2 te berekenen. Het eigenvector probleem kan op een analoge wijze opgelost worden. Als bv. $B_1\hat{x} = \lambda\hat{x}$ met $\hat{x} \in \mathrm{R}^2 \setminus [0,0]^T$, definieer dan

$$x = [\hat{x}^T, 0, 0, 0]^T \,,$$

en dan is $Bx = \lambda x$. Die wijze van werken is bruikbaar voor de bepaling van het compleet stel eigenvectoren van B uit deze van B_1 en B_2 .

In de verdere paragrafen onderstellen we dat alle $\beta_l \neq 0$ in de matrix B .

4.3.5 De Sturm rij eigenschappen van $\{f_k(\lambda)\}$

De rijen $\{f_k(a)\}$ en $\{f_k(b)\}$ kunnen gebruikt worden om het aantal wortels van $f_n(\lambda)$ te bepalen die gelegen zijn in $[a, b]$. Om dit te verwezenlijken introduceren we de gehele functie $s_k(\lambda)$, die het aantal overeenstemmingen in teken aangeeft van opeenvolgende leden van de rij $\{f_0(\lambda), f_1(\lambda), \ldots, f_k(\lambda)\}$. Als de waarde van een bepaald lid bv. $f_j(\lambda) = 0$, dan wordt zijn teken tegengesteld gekozen aan dat van $f_{j-1}(\lambda)$. We zullen later tonen dat $f_j(\lambda) = 0$ impliceert dat $f_{j-1}(\lambda) \neq 0$.

Voorbeeld 4.3.1

Beschouw

$$B = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \tag{4.31}$$

en $f_0(\lambda) = 1\,,\ f_1(\lambda) = 2 - \lambda$
en (vergelijk met (4.28) en (4.30))

$$f_j(\lambda) = (2 - \lambda)f_{j-1}(\lambda) - f_{j-2}(\lambda) \qquad j = 2, 3, 4, 5, 6 \tag{4.32}$$

Voor $\lambda = 3$ krijgen we

$$f_0(3) = 1\,,\ f_1(3) = -1\,,\ f_2(3) = 0\,,\ f_3(3) = 1\,,\ f_4(3) = -1\,,$$
$$f_5(3) = 0\,,\ f_6(3) = 1\,.$$

Dus de corresponderende tekenrij is

$$(+, -, +, +, -, +, +)$$

en volgens bovenstaande afspraak is $s_6(3) = 2$. ◇

Theorema 4.3.1

Zij B een reële symmetrische tridiagonale matrix van orde n zoals gegeven in (4.28). Zij de rij $\{f_k(\lambda) \mid 0 \leq k \leq n\}$ gedefinieerd zoals in (4.30) en weze alle $\beta_l \neq 0$ $(l = 1, \ldots, n-1)$. Het aantal wortels van $f_n(\lambda)$ dat groter is dan $\lambda = a$ wordt dan gegeven door $s_n(a)$, waarbij $s_n(\lambda)$ gedefinieerd is in het begin van deze paragraaf. Voor $a < b$ wordt het aantal wortels in het interval $a < \lambda \leq b$ gegeven door $s_n(a) - s_n(b)$.

Bewijs

Het bewijs van bovenstaand theorema is nogal lang en is opgesplitst in verscheidene delen.

(1) Twee opeenvolgende polynomen uit de rij hebben geen gemeenschappelijke wortel. Voor een bewijs uit het ongerijmde veronderstel dat

$$f_j(\lambda) = f_{j-1}(\lambda) = 0 \quad \text{voor een bepaalde } j \geq 2.$$

Dan volgt uit (4.30) met $k = j$

$$f_{j-2}(\lambda) = \frac{1}{\beta_{j-1}^2}\Big[(\alpha_j - \lambda) f_{j-1}(\lambda) - f_j(\lambda)\Big] = 0\,.$$

Deze argumentatie verder zettend voor alle $k \leq j$ leidt tot

$$f_k(\lambda) = 0 \qquad \text{voor alle} \qquad k \leq j$$

wat in strijd is met de definitie $f_0(\lambda) \equiv 1$.

(2) De wortels van $f_k(\lambda)$ worden strikt gescheiden door die van $f_{k-1}(\lambda)$ voor $k = 2, 3, \ldots, n$. Voor het eenvoudigste geval beschouw de wortels van $f_1(\lambda)$ en $f_2(\lambda)$. De enkelvoudige wortel van $f_1(\lambda)$ is $\lambda = \alpha_1$. Vermits $f_2(\pm\infty) = +\infty$ en $f_2(\alpha_1) = -\beta_1^2 < 0$ ligt er aan elke kant van $\lambda = \alpha_1$ één wortel van $f_2(\lambda)$.
Neem nu aan dat het resultaat waar is voor de wortels van $f_1(\lambda), \ldots, f_{k-1}(\lambda)$. We zullen nu bewijzen dat bovenstaande bewering ook correct is voor de wortels van $f_k(\lambda)$. Noteren we de wortels van $f_{k-1}(\lambda)$ en $f_{k-2}(\lambda)$ als $\{\lambda_1, \ldots, \lambda_{k-1}\}$ en $\{\mu_1, \ldots, \mu_{k-2}\}$ respectievelijk. Hiervoor geldt bij veronderstelling

$$\lambda_{k-1} < \mu_{k-2} < \lambda_{k-2} < \ldots < \lambda_2 < \mu_1 < \lambda_1\,. \tag{4.33}$$

Noteer dat, vermits $f_{k-1}(\lambda)$ van graad $k-1$ en vermits er exact $k-1$ verschillende wortels $\lambda_1, \ldots, \lambda_{k-1}$ zijn, al deze wortels enkelvoudig moeten zijn. Hetzelfde is waar voor de wortels van $f_{k-2}(\lambda)$. Laten we nu het teken bepalen van $f_k(\lambda)$ bij $\lambda = \lambda_{j-1}$ en $\lambda = \lambda_j$. Uit (4.30) volgt :

$$\begin{aligned} f_k(\lambda_j) &= (\alpha_k - \lambda_j) f_{k-1}(\lambda_j) - \beta_{k-1}^2 f_{k-2}(\lambda_j) \\ &= -\beta_{k-1}^2\, f_{k-2}(\lambda_j) \end{aligned}$$

en

$$f_k(\lambda_{j-1}) = -\beta_{k-1}^2 \, f_{k-2}(\lambda_{j-1}) \, . \tag{4.34}$$

Vermits $f_{k-2}(\lambda)$ een enkelvoudig wortel μ_{j-1} bezit tussen λ_{j-1} en λ_j is het nodig dat

$$\text{teken}(f_{k-2}(\lambda_j)) = -\text{teken}(f_{k-2}(\lambda_{j-1})) \, .$$

Maar uit (4.34) volgt dat hetzelfde moet gelden voor $f_k(\lambda_j)$ en $f_k(\lambda_{j-1})$ en daarom moet $f_k(\lambda)$ een wortel hebben van oneven multipliciteit tussen λ_{j-1} en λ_j. Dit is correct voor elke $j = 2, 3, \ldots, k-1$.
Vermits nu

$$f_l(\lambda) = (-1)^l \lambda^l \, + \, \text{lagere graads termen} \quad (\text{voor alle } l \geq 1)$$

hebben we dat

$$\begin{aligned} \text{teken}\,(f_k(-\infty)) &= \text{teken}\,(f_{k-2}(-\infty)) \\ \text{teken}\,(f_k(+\infty)) &= \text{teken}\,(f_{k-2}(+\infty)). \end{aligned} \tag{4.35}$$

Tevens volgt uit (4.34)

$$\text{teken}\,(f_k(\lambda_1)) = -\text{teken}\,(f_{k-2}(\lambda_1)) = -\text{teken}\,(f_{k-2}(+\infty)) \, . \tag{4.36}$$

De laatste gelijkheid volgt uit het feit dat f_{k-2} geen wortels rechts van μ_1 bezit. Uit (4.35) en (4.36) volgt dat $f_k(\lambda)$ een wortel rechts van λ_1 moet bezitten. Een analoge redenering toont aan dat $f_k(\lambda)$ een wortel links van λ_{k-1} moet bezitten. Wanneer we nu de gelocaliseerde wortels van $f_k(\lambda)$ tellen, komen we aan k, wat precies in overeenstemming is met graad $(f_k(\lambda)) = k$. Dit impliceert dat alle wortels enkelvoudig zijn en de wortels van $f_{k-1}(\lambda)$ inderdaad deze van $f_k(\lambda)$ scheiden.

(3) Het aantal wortels van $f_k(\lambda)$ welke groter zijn dan a wordt gegeven door $s_k(a)$. We zullen het bewijs leveren door inductie.
Voor de rij met $k = 1$

$$\{f_0(a) \, , \, f_1(a)\} = \{1, \, \alpha_1 - a\}$$

is het duidelijk dat het aantal overeenkomsten in teken gelijk is aan het aantal wortels van $f_1(\lambda)$ groter dan a. Neem nu aan dat het gestelde waar is voor de rij

$$f_0(a) \, , \ldots , \, f_{k-1}(a)$$

en laten we het aantal wortels van $f_{k-1}(\lambda)$ die groter zijn dan a, noteren als m. Zij $\{\lambda_i\}$, respectievelijk $\{\nu_i\}$ de verzameling wortels van $f_{k-1}(\lambda)$, respectievelijk $f_k(\lambda)$ dan impliceren de zo juist gemaakte hypothese en de resultaten uit (2) dat

$$\lambda_1 > \lambda_2 > \ldots \lambda_m > a \geq \lambda_{m+1} > \ldots > \lambda_{k-1} \tag{4.37}$$

$$\nu_1 > \lambda_1 > \nu_2 > \ldots > \nu_m > \lambda_m > \nu_{m+1} > \lambda_{m+1} > \ldots > \lambda_{k-1} > \nu_k \,. \tag{4.38}$$

We wensen nu te bewijzen dat het aantal wortels van $f_k(\lambda)$ die groter zijn dan a, gelijk is aan het aantal overeenkomsten in teken in de rij

$$\{f_0(a)\,, \ldots,\, f_k(a)\}. \tag{4.39}$$

Uit (4.37) en (4.38) volgt dat het aantal wortels van $f_k(\lambda)$ dat groter is dan a, hetzij m, hetzij $m+1$ is. Het bewijs van ons gewenst resultaat breekt op in drie gevallen :

(a) $a \notin \{\lambda_{m+1}\,,\, \nu_{m+1}\}$;

Schrijf hiertoe

$$f_{k-1}(\lambda) = (\lambda_1 - \lambda) \ldots (\lambda_{k-1} - \lambda) \tag{4.40}$$

$$f_k(\lambda) = (\nu_1 - \lambda) \ldots (\nu_k - \lambda)\,. \tag{4.41}$$

Hierin zijn twee bijzondere gevallen te beschouwen.
Vooreerst als $\nu_{m+1} > a > \lambda_{m+1}$ dan is het aantal wortels van $f_k(\lambda)$ groter dan a gelijk aan $m+1$. Uit (4.41) volgt ook

$$\text{teken}\,(f_{k-1}(a)) = (-1)^{(k-1)-m} \quad \text{en} \quad \text{teken}\,(f_k(a)) = (-1)^{k-(m+1)}.$$

Vermits die tekens gelijk zijn is het aantal overeenkomsten in teken in de rij (4.39) $m+1$, wanneer we gebruik maken van het feit dat initiaal het aantal m was voor de rij

$$\{f_0(a), \ldots,\, f_{k-1}(a)\}\,.$$

Voor het tweede subgeval, $\lambda_m > a > \nu_{m+1}$ is het aantal wortels van $f_k(\lambda)$ groter dan a, juist m. Evenzo zijn

$$\text{teken}\,(f_{k-1}(a)) = (-1)^{(k-1)-m} \quad \text{en} \quad \text{teken}\,(f_k(a)) = (-1)^{k-m}$$

verschillend .

(b) $a = \nu_{m+1}$;

Dan is $f_k(a) = 0$ en dan is bij afspraak het teken van $f_k(a)$ tegengesteld aan dit van $f_{k-1}(a)$, d.w.z. dat het aantal overeenkomsten in teken in (4.39) m blijft, hetzelfde als het aantal wortels van $f_k(\lambda)$ die groter zijn dan $a = \nu_{m+1}$;

(c) $a = \lambda_{m+1}$;

Dan is $f_{k-1}(a) = 0$ en er zijn $m+1$ wortels van $f_k(\lambda)$ groter dan a. Uit (4.30) volgt er

$$f_k(a) = -\beta_{k-1}^2\, f_{k-2}(a)$$

en gelet op de ingevoerde teken conventie

$$\text{teken}\,(f_{k-1}(a)) = -\text{teken}\,(f_{k-2}(a))\,.$$

Beide voorgaande relaties combinerend, levert

$$\text{teken}\,(f_k(a)) = \text{teken}\,(f_{k-1}(a))$$

en dus is het aantal overeenkomstige tekens in (4.39) $m+1$, d.i. het gewenste resultaat.

Dit vervolledigt het bewijs. $\diamond$

4.3.6 Bepaling van de eigenwaarden van een tridiagonale matrix door gebruik te maken van de eigenschappen van Sturm rijen

Het theorema uit voorgaande subparagraaf zal aangewend worden als basismiddel om de wortels van $f_n(\lambda)$ te localiseren en te scheiden. Om te beginnen omschrijven we een interval dat al de wortels bevat. Hiertoe steunen we op het Gershgorin cirkeltheorema.

Theorema 4.3.2 (Gershgorin cirkeltheorema)

Het spectrum van een $n \times n$ matrix A (dat is, de verzameling van zijn eigenwaarden) is een deelverzameling van de unie van de volgende n schijven in het complexe vlak :

$$D_i = \left\{ z \in \mathrm{C} : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right\} \qquad (1 \leq i \leq n)\,.$$

Bewijs

Weze λ een willekeurig element van het spectrum van A. Selecteer dan een vector x zodat $Ax = \lambda x$ en $\|x\|_\infty = 1$. Zij i een index waarvoor $|x_i| = 1$. Vermits $(Ax)_i = \lambda x_i$ hebben we dat

$$\lambda x_i = \sum_{j=1}^{n} a_{ij} x_j\,.$$

Daarom is

$$(\lambda - a_{ii}) x_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j\,.$$

Door van linker- en rechterlid de absolute waarde te nemen en rekening te houden met de driehoeksongelijkheid en het feit dat $|x_j| \leq 1 = |x_i|$, resulteert er dat

$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}||x_j| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \,.$$

Dus $\lambda \in D_i$, d.w.z. als λ een eigenwaarde van A is, dan behoort λ tot tenminste één van de schijven D_i. $\diamond$

Toegepast nu op reële matrices zijn uiteraard alle a_{ii} reëel en kunnen we in het geval van de tridiagonale matrix B voorspellen dat alle eigenwaarden liggen in het interval $[a, b]$ met

$$\begin{aligned} a &= \min_{1 \leq i \leq n} \Big\{ \alpha_i - \mid \beta_i \mid - \mid \beta_{i-1} \mid \Big\} \\ b &= \max_{1 \leq i \leq n} \Big\{ \alpha_i + \mid \beta_i \mid + \mid \beta_{i-1} \mid \Big\} \end{aligned}$$

en met $\beta_0 = \beta_n = 0$.
We kunnen dan het interval $[a, b]$ in kleinere subintervallen opsplitsen. Het theorema aangaande de Sturm rijen (zie par. 4.3.5) kan toegepast worden om het aantal wortels (dit zijn de eigenwaarden), gelegen in een subinterval, te bepalen. We streven ernaar finaal subintervallen aan te duiden die slechts één wortel bevatten. Eens dit doel bereikt, kunnen we elke wortel met een sneller iteratiemethode exacter bepalen.

Voorbeeld 4.3.2

Beschouwen we opnieuw het voorbeeld (4.31). Steunend op het Gershgorin theorema kunnen we vooropstellen dat alle eigenwaarden liggen in het interval $[0, 4]$. Een systematisch halveringsproces kan uitgevoerd worden op $[0, 4]$ om de zes wortels van $f_6(\lambda)$ in te delen in zes verschillende subintervallen. Het resultaat is opgesomd in onderstaande tabel. De eigenwaarden worden op de volgende wijze genoteerd :

$$0 \leq \lambda_6 < \lambda_5 < \ldots < \lambda_1 \leq 4$$

λ	$f_0(\lambda)$	$f_1(\lambda)$	$f_2(\lambda)$	$f_3(\lambda)$	$f_4(\lambda)$	$f_5(\lambda)$	$f_6(\lambda)$	$s_6(\lambda)$	kommentaar
0	1	2	3	4	5	6	7	6	$0 < \lambda_6$
4	1	-2	3	-4	5	-6	7	0	$\lambda_1 \leq 4$
2	1	0	-1	0	1	0	-1	3	$\lambda_4 \leq 2 < \lambda_3$
1	1	1	0	-1	-1	0	$+1$	4	$\lambda_5 \leq 1 < \lambda_4 \leq 2$
0.5	1	1.5	1.25	0.375	-0.6875	-1.40625	-1.421875	5	$0 < \lambda_6 \leq 0.5 < \lambda_5 \leq 1$
3	1	-1	0	1	-1	0	1	2	$2 < \lambda_3 \leq 3 < \lambda_2$
3.5	1	-1.5	1.25	-0.375	-0.6875	1.40625	-1.412875	1	$3 < \lambda_2 \leq 3.5 < \lambda_1 \leq 4$

Exactere waarden voor elk van deze wortels kunnen in elk subinterval bekomen worden met bijvoorbeeld de halverings–, de Regula-falsi– of de Newton-Raphson–methode. Deze manier van werken biedt de mogelijkheid alleen een beperkt aantal eigenwaarden, bv. de grootste of de kleinste te bepalen, wat bijvoorbeeld met de Jacobi–methode niet het geval was. $\diamond$

4.3.7 Eigenvectoren van tridiagonale matrices

Wanneer λ_1 een exacte eigenwaarde is van de matrix B (4.28) dan moeten we zoeken naar een vector x, die een oplossing is van $Bx = \lambda_1 x$. Vermits dit een homogeen stelsel is in n veranderlijken, en vermits $\det[B - \lambda_1 I] = 0$, kunnen we een niet–triviale oplossing verkrijgen door $(n-1)$ vergelijkingen te kiezen en hieruit de componenten van x op een constante factor na te bepalen; de overblijvende vergelijking hoeft dan uiteraard automatisch voldaan te zijn. In de praktijk blijkt dat zelfs voor goed geconditioneerde matrices de nauwkeurigheid van het resultaat in grote mate bepaald wordt door de vergelijking die vooraf uitgesloten werd. Laat ons bijvoorbeeld aannemen dat de i^{de} vergelijking niet in acht werd genomen, terwijl de andere opgelost werden door eliminatie. De oplossing (exact verondersteld) voldoet aan de $(n-1)$ vergelijkingen die gebruikt werden voor de eliminatie, maar levert een fout δ wanneer ze ingevoerd wordt in de i^{de} vergelijking. In feite kan de berekende vector x opgevat worden als de oplossing van het stelsel

$$\begin{cases} \beta_{j-1}\, x_{j-1} + (\alpha_j - \lambda) x_j + \beta_j\, x_{j+1} = 0 & \qquad j \neq i \\ \beta_{i-1}\, x_{i-1} + (\alpha_i - \lambda) x_i + \beta_i\, x_{i+1} = \delta & \qquad \delta \neq 0 \end{cases}$$

Hierin betekent λ een vooraf berekende benadering voor de exacte eigenwaarde λ_1. Vermits constante factoren hier toch geen rol spelen, kan men $\delta = 1$ stellen en kan bovenstaand stelsel eenvoudiger genoteerd worden als

$$(B - \lambda I)x = e_i \tag{4.42}$$

waarin e_i de kolomvector is met de i^{de} component gelijk aan 1 en de overige componenten gelijk aan nul. Als de exacte eigenvectoren van B, $v_1,\ v_2, \ldots,\ v_n$ zijn, dan kan de vector e_i uitgedrukt worden als een lineaire combinatie ervan. Immers deze eigenvectoren vormen een compleet stel lineair onafhankelijke vectoren, m.a.w. ze vormen een basis en elke n–dimensionale vector kan neergeschreven worden als een lineaire combinatie van basis vectoren, d.i.

$$e_i = \sum_{j=1}^{n} c_{ij}\, v_j \tag{4.43}$$

en wegens (4.42)

$$x = \sum_{j=1}^{n} c_{ij} (B - \lambda I)^{-1} v_j = \sum_{j=1}^{n} c_{ij} \frac{1}{\lambda_j - \lambda}\, v_j \ .$$

Als nu $\lambda = \lambda_1 + \epsilon$, herleidt bovenstaande betrekking zich tot

$$x = -\frac{c_{i1}}{\epsilon}\, v_1 + \sum_{j=2}^{n} c_{ij} \frac{1}{\lambda_j - \lambda_1 - \epsilon}\, v_j \ . \tag{4.44}$$

Wanneer $c_{i1} \neq 0$ nadert de oplossing x naar v_1 als $\epsilon \to 0$. Het kan echter ook gebeuren dat c_{i1} van dezelfde orde van grootte is als ϵ. Uit (4.43) kunnen we dan afleiden dat e_i bijna orthogonaal is t.o.v. v_1. Indien dit zo is, dan kan de vector x in (4.44) geen goede benadering voor v_1 zijn. Om nooit met dit probleem geconfronteerd te worden, suggereert Wilkinson (J. Wilkinson, *The Algebraic Eigenvalue Problem.* Oxford University Press, Oxford, England 1965) (4.42) te vervangen door

$$(B - \lambda I)x = b$$

waarbij we de vector b vrij tot onze beschikking hebben. Dit stelsel kan dan opgelost worden door een Gauss eliminatiemethode, waarbij er gewaakt moet worden over het feit dat de vergelijkingen op een degelijke wijze gepermuteerd worden om het pivot element zo groot mogelijk te maken. Het resulterende stelsel kan dan als volgt genoteerd worden :

$$\left\{\begin{array}{rcrcrcrcl} p_{11}x_1 & + & p_{12}x_2 & + & p_{13}x_3 & & & = & c_1 \\ & & p_{22}x_2 & + & p_{23}x_3 & + & p_{24}x_4 & = & c_2 \\ & & & & & & & \vdots & \\ & & & & & & p_{nn}x_n & = & c_n. \end{array}\right. \tag{4.45}$$

De meeste van de coëfficiënten $p_{13},\ p_{24} \ldots$ zullen nul zijn. Vermits de c_i bekomen worden uit de b_i die we tot onze beschikking hebben, kunnen we dus even goed de constanten c_i vrij kiezen. Een aanvaardbare keuze bestaat erin alle c_i gelijk 1 te kiezen, i.e.

$$c = \sum_{r=1}^{n} e_r \ .$$

Het stelsel (4.45) wordt dan van boven naar onder opgelost en de vector x wordt tenslotte genormeerd.

4.4 De machtmethode (power method)

Voor tekst uitleg en implementatie van die methode zie derde oefeningsessie Matlab

---

Algemene nota

In de voorgaande paragrafen is gebleken dat de oplossing van eigenwaarde problemen een zeer ingewikkelde problematiek is. Het is één van de onderwerpen uit het numeriek rekenen waar, bij praktisch gebruik, het zeer wenselijk is beschikbare routines uit bibliotheken toe te passen. De meeste aangeboden routines in pakketten vinden hun wortels in Wilkinson and Reinsch, *Handbook for Automatic Computation, Vol. II, Linear Algebra.* Dit excellent referentiewerk, dat artikels bevat van een aantal auteurs, is de bijbel van dit onderzoeksgebied. Een publiek toegankelijke implementatie van de "Handbook" routines in FORTRAN is

de EISPACK verzameling van programma's (zie B.T. Smith et al., *Matrix Eigensystem Routines – EISPACK Guide, 2nd ed., vol. 6 of Lecture Notes in Computer Science*). Dergelijke routines worden ook geleverd door *NAG Fortran Library Manual* en kan men ook vinden in W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vettering, *Numerical Recipes, The Art of Scientific Computing* met uitgewerkte programma's in Pascal, FORTRAN en C. Ook in de NAG-toolbox van Matlab zijn verschillende functies aanwezig om eigenwaarden en eigenvectoren te bepalen (zie derde oefeningsessie). Elk goed "eigenwaarde pakket" moet voorzien in afzonderlijke procedures voor de volgende mogelijk te verlangen berekeningen :

– eigenwaarden en geen eigenvectoren
– alle eigenwaarden en sommige corresponderende eigenvectoren
– alle eigenwaarden en alle eigenvectoren.

Het doel van deze verschillen is computertijd en stockeringsruimte te besparen. Een goed pakket voorziet ook in speciale mogelijkheden van de bovengenoemde grootheden voor elk van de volgende speciale vormen van de matrix :

– reëel, symmetrisch, tridiagonaal
– reëel, symmetrisch met bandstructuur
– reëel, symmetrisch
– reëel, niet-symmetrisch
– complex, Hermitisch
– complex, niet-Hermitisch.

De algoritmen voor symmetrisch matrices, besproken in dit hoofdstuk, geven voldoening in de praktijk. Het is echter niet mogelijk even goede algoritmen voor niet-symmetrische matrices te ontwikkelen. Hiervoor zijn twee redenen. Vooreerst blijken de eigenwaarden van niet-symmetrische matrices zeer gevoelig te zijn voor kleine veranderingen in de waarden van de matrixelementen. In de tweede plaats bezitten niet-symmetrische matrices niet altijd een complete verzameling eigenvectoren. De algemene strategie voor het vinden van eigenwaarden voor dergelijke matrices bestaat erin de matrix om te vormen tot een zgn. *Hessenberg vorm*. Een boven Hessenberg matrix heeft onder de diagonaal allemaal nul elementen, behalve op de eerste subdiagonaal. Nadien wordt dan meestal het zgn. QR-algoritme van Francis toegepast om de eigenwaarden (reëel of complex) te isoleren. De volledige uiteenzetting van deze techniek valt buiten het kader van deze cursus.

De methode van Faddeev-Leverrier maakte deel uit van de technieken en numerieke methoden waarmee Leverrier in 1845 de plaats en het bestaan van de planeet Neptunus voorspelde. *Charles Sturm* (1803–1855) is een Frans wiskundige van Zwitserse afkomst, vooral bekend voor de stelling die zijn naam draagt en die toelaat het aantal reële wortels van een hogere machtsvergelijking, gelegen in een voorgeschreven interval, te bepalen. De machtmethode wordt dikwijls toegeschreven aan *Richard Edler von Mises* (1883–1953), die werkte in Wenen, Brno, Straatsburg, Dresden, Berlijn, Istanboel en Harvard University. Zijn wiskundige belangstelling vond toepassingen zowel in de mechanica als in de waarschijnlijkheidsrekening.