Nove metode za računanje spektralne dekompozicije simetrične matrice reda 3 i 4

Kranjčević, Marija

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Marija Kranjčević

NOVE METODE ZA RAČUNANJE SPEKTRALNE DEKOMPOZICIJE SIMETRIČNE MATRICE REDA 3 I 4

Diplomski rad

Voditelj rada:
prof.dr.sc. Vjeran Hari

Zagreb, 2015.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Uvod

Cilj ovog rada je opisati i implementirati nove metode za računanje spektralne dekompozicije simetričnih matrica malog reda. Ti algoritmi se zatim mogu koristiti kao jezgreni algoritmi za računanje blok dijagonalizacijskih algoritama za simetrične matrice reda n. Mi ćemo opisati, analizirati i implementirati algoritam za računanje spektralne dekompozicije simetričnih matrica reda 3 i 4 koji se bazira na algoritmu za spektralnu dekompoziciju streličastih matrica zvanom *aheig* (ArrowHEead EIGenvalues/eigenvectors, [7, 8]).

U prvom, uvodnom poglavlju, uvodimo osnovne oznake, te navodimo temeljne definicije i teoreme bitne za računanje spektralne dekompozicije simetričnih matrica.

U drugom poglavlju definiramo simetričnu ireducibilnu streličastu matricu, opisujemo algoritam *aheig* za računanje njene spektralne dekompozicije, te dokazujemo da on, pod određenim uvjetima, sve svojstvene vrijednosti i sve komponente pripadnih svojstvenih vektora računa s visokom relativnom točnošću. U drugom poglavlju zbog preglednosti navodimo samo tvrdnje, a njihove dokaze dajemo u Dodatku. Opisane algoritme implemetiramo u Matlabu, a analizu točnosti i tok algoritma ilustriramo na konkretnim primjerima.

U trećem poglavlju opisujemo Jacobijevu rotaciju, dajemo neke rezultate o njenoj točnosti, te implementaciju u Matlabu.

U četvrtom poglavlju kombiniramo Jacobijevu rotaciju s algoritmom *aheig* kako bi dobili spektralnu dekompoziciju simetrične matrice reda 3, a zatim, korištenjem opisanog postupka, i spektralnu dekompoziciju simetrične matrice reda 4. Primjenom Jacobijeve rotacije od proizvoljne simetrične matrice reda 3 dobivamo streličastu matricu, čiju spektralnu dekompoziciju onda možemo izračunati algoritmom *aheig*. U slučaju simetrične matrice reda 4, prvo na opisan način izračunamo spektralnu dekompoziciju njene podmatrice, čime dobivamo streličastu matricu reda 4, a zatim ponovo možemo primijeniti algoritam *aheig*.

Poglavlje 1

Osnovni pojmovi, oznake i rezultati

U ovom poglavlju uvodimo osnovne oznake, te navodimo temeljne definicije i teoreme bitne za računanje spektralne dekompozicije simetričnih matrica.

1.1 Oznake

Polje realnih brojeva označavamo s $\mathbb{R}$, a skalarne veličine uglavnom malim grčkim slovima. Vektorski prostor uređenih n-torki realnih brojeva označavamo s $\mathbb{R}^n$, a vektorski prostor realnih matrica tipa $m \times n$ s $\mathbb{R}^{m\times n}$. Elemente prostora R^n zovemo vektori i označavamo malim latinskim slovima, a matrice označavamo velikim latinskim slovima. Ako je $\mathcal{X}$ potprostor od $\mathbb{R}$, pišemo $\mathcal{X} \leq \mathbb{R}$, a njegov ortogonalni komplement označavamo s $\mathcal{X}^\perp$.
Neka je $x \in \mathbb{R}^n$. Tada pišemo $x = [x_i]_{i=1}^n$, $x = [x_i]$ ili

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

tj. i-ta komponenta vektora je jednaka x_i. Za $A \in \mathbb{R}^{m\times n}$ pišemo $A = [a_{ij}]$ i

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix},$$

tj. element matrice na mjestu (i, j) je a_{ij}.

Koristimo i Matlab-ove oznake

$$x(i) = x_i,$$
$$x(i:j) = \begin{bmatrix} x_i \\ \vdots \\ x_j \end{bmatrix},$$
$$A(i,j) = a_{ij},$$
$$A(i:j, k:l) = \begin{bmatrix} a_{ik} & \dots & a_{il} \\ \vdots & & \vdots \\ a_{jk} & \dots & a_{jl} \end{bmatrix},$$
$$A(:, k:l) = A(1:m, k:l),$$
$$A(i:j, :) = A(i:j, 1:n).$$

S $I_n = [e_1 \dots e_n]$ označavamo jediničnu matricu, a s $0_{(m\times n)}$ nul-matricu reda $m \times n$. S A^T i A^{-1} označavamo transponiranu i inverznu matricu matrice A, respektivno. Kažemo da je matrica $A \in \mathbb{R}^{n\times n}$ simetrična ako vrijedi $A^T = A$, ortogonalna ako je $A^T A = I_n$, te dijagonalna ako je $a_{ij} = 0$, za sve $i \neq j,\ i, j \in \{1, \dots, n\}$. Koristimo i standardni skalarni produkt u $\mathbb{R}^n$

$$\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i,$$

vektorske norme $\|x\|_2 = \sqrt{x^T x},\ \|x\|_\infty = \max\limits_{1\le i\le n} |x_i|$, te matrične norme

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2,$$
$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|,$$
$$\|A\|_F = \sqrt{\operatorname{trag}(A^T A)} = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2}.$$

Za egzaktnu vrijednost x izračunatu vrijednost označavamo s $\widetilde{x} = fl(x)$. Egzaktne vrijednosti ponekad označavamo i s $\hat{}$, a jediničnu grešku zaokruživanja računala s ε_M. Pretpostavljamo da računamo u aritmetici s korektnim zaokruživanjem tj. da vrijedi

$$fl(x \circ y) = (x \circ y)(1 + \varepsilon), \quad |\varepsilon| \le \varepsilon_M, \quad \circ \in \{+, -, *, /, \sqrt{\cdot}\},$$

gdje su x i y reprezentabilni brojevi. Pretpostavljamo i da nema prekoračenja ni potkoračenja. Također, u analizi grešaka zaokruživanja zanemarujemo članove reda veličine $O(\varepsilon_M^k)$, $k \ge 2$.

1.2 Svojstvene vrijednosti i svojstveni vektori

Neka je $A \in \mathbb{R}^{n\times n}$. Ako vrijedi

$$Ax = \lambda x,\ x \neq 0, \quad \lambda \in \mathbb{R},\ x \in \mathbb{R}^n,$$

onda kažemo da je λ svojstvena vrijednost, a x svojstveni vektor matrice A. Za svaki skalar $\alpha \neq 0$, vektor αx je također svojstveni vektor od A pridružen svojstvenoj vrijednosti λ. Uređeni par (λ, x) zovemo svojstveni par od A. Skup svih svojstvenih vrijednosti nazivamo spektar matrice A i označavamo sa $\sigma(A)$. Nadalje pretpostavljamo da je $A \in \mathbb{R}^{n\times n}$ simetrična matrica. Tada je njena spektralna dekompozicija

$$A = U\Lambda U^T, \tag{1.1}$$

gdje je $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_1 \geq \cdots \geq \lambda_n$ svojstvene vrijednosti u padajućem poretku, a $U = [u_1\ \ldots\ u_n]$ orotgonalna matrica. Posebno, $u_1, \ldots, u_n$ su normirani svojstveni vektori tj. vrijedi $\|u_i\|_2 = 1$, za $i = 1, \ldots, n$.

Slijede osnovne tvrdnje teorije spektralne dekompozicije i simetričnih matrica, koje ćemo kasnije koristiti.

Teorem 1.1 (Geršgorin, [2, 11])**.** *Neka je $A \in \mathbb{R}^{n\times n}$ simetrična matrica. Tada vrijedi*

$$\sigma(A) \subseteq \bigcup_{i=1}^{n} [a_{ii} - r_i, a_{ii} + r_i], \quad \textit{gdje je } r_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|,\ i = 1 \ldots, n.$$

Dokaz. Neka je (1.1) spektralna dekompozicija matrice A i neka je $i \in \{1, \ldots, n\}$. Tada imamo

$$Au_i = \lambda_i u_i, \tag{1.2}$$

gdje je $u_i = [u_1^{(i)}, \ldots, u_n^{(i)}]^T$. Neka je $j \in \{1, \ldots, n\}$ takav da je $|u_j^{(i)}| = \|u_i\|_\infty$. Tada j-ta jednadžba u (1.2) glasi

$$\sum_{k=1}^{n} a_{jk} u_k^{(i)} = \lambda_i u_j^{(i)}. \tag{1.3}$$

Budući da je u_i svojstveni vektor, vrijedi $u_i \neq 0$. Zato je $|u_j^{(i)}| = \|u_i\|_\infty \neq 0$, pa onda i $u_j^{(i)} \neq 0$. Dijeljenjem jednadžbe (1.3) s $u_j^{(i)}$ dobivamo

$$\sum_{\substack{k=1 \\ k\neq j}}^{n} a_{jk} \frac{u_k^{(i)}}{u_j^{(i)}} = \lambda_i - a_{jj}.$$

Slijedi

$$|\lambda_i - a_{jj}| \leq \sum_{\substack{k=1 \\ k\neq j}}^{n} |a_{jk}| \underbrace{\left|\frac{u_k^{(i)}}{u_j^{(i)}}\right|}_{\leq 1} \leq \sum_{\substack{k=1 \\ k\neq j}}^{n} |a_{jk}| = r_j.$$

Dakle, $\lambda_i \in [a_{jj} - r_j, a_{jj} + r_j]$, pa vrijedi i $\lambda_i \in \bigcup_{k=1}^{n} [a_{kk} - r_k, a_{kk} + r_k]$. Budući da je $i \in \{1, \ldots, n\}$ bio proizvoljan, slijedi tvrdnja. □

Lema 1.2 ([11])**.** *Neka je $A \in \mathbb{R}^{n\times n}$ simetrična matrica i neka je s (1.1) dana njena spektralna dekompozicija. Tada za $k = 1, \ldots, n$ vrijedi*

$$\lambda_k = \max_{\substack{x^T x=1 \\ x\perp u_1,\ldots,u_{k-1}}} x^T A x = \min_{\substack{x^T x=1 \\ x\perp u_{k+1},\ldots,u_n}} x^T A x.$$

Dokaz. Uzmimo normirani vektor $x \in \mathbb{R}^n$ tj. neka vrijedi $\|x\|_2^2 = x^T x = 1$. Uz $y = U^T x$ vrijedi

$$x^T A x = x^T U \Lambda U^T x = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2 \in [\lambda_n, \lambda_1], \tag{1.4}$$

jer je $\sum_{i=1}^{n} y_i^2 = y^T y = x^T U U^T x = x^T x = 1$. Odmah vidimo da je

$$\lambda_n = u_n^T A u_n = \min_{x^T x=1} x^T A x,$$
$$\lambda_1 = u_1^T A u_1 = \max_{x^T x=1} x^T A x.$$

Nadalje, iz (1.4) vidimo da uvjet $x \perp u_1$ (tj. $u_1^T x = 0$) daje $y_1 = 0$, pa je

$$x^T A x = \sum_{i=2}^{n} \lambda_i y_i^2 \in [\lambda_n, \lambda_2]$$

i

$$\lambda_2 = u_2^T A u_2 = \max_{\substack{x^T x=1 \\ x\perp u_1}} x^T A x.$$

Analogno, uzimanjem $x \perp u_n$, dobivamo

$$\lambda_{n-1} = u_{n-1}^T A u_{n-1} = \min_{\substack{x^T x=1 \\ x\perp u_n}} x^T A x.$$

Dodavanjem novih uvjeta ortogonalnosti dobivamo da za $k = 1, \ldots, n$ vrijedi

$$\lambda_k = u_k^T A u_k = \max_{\substack{x^T x = 1 \\ x \perp u_1, \ldots, u_{k-1}}} x^T A x = \min_{\substack{x^T x = 1 \\ x \perp u_{k+1}, \ldots, u_n}} x^T A x.$$

□

Teorem 1.3 (Cauchyjev teorem o preplitanju, [9]). *Neka je $A \in \mathbb{R}^{n \times n}$ simetrična matrica i neka je (1.1) njena spektralna dekompozicija. Označimo*

$$A = \begin{bmatrix} B & c \\ c^T & d \end{bmatrix}, \ c \in \mathbb{R}^{n-1}, \ d \in \mathbb{R},$$

te neka je

$$B y_i = \mu_i y_i, \quad \|y_i\|_2 = 1, \quad i = 1, \ldots, n-1, \quad \mu_1 \geq \cdots \geq \mu_{n-1}.$$

Tada vrijedi

$$\lambda_k \geq \mu_k \geq \lambda_{k+1}, \quad k = 1, \ldots, n-1.$$

Dokaz. Iz (1.1) imamo

$$A u_i = \lambda_i u_i, \quad \|u_i\|_2 = 1, \quad i = 1, \ldots, n, \quad \lambda_1 \geq \cdots \geq \lambda_n.$$

Neka je $k \in \{1, \ldots, n-1\}$ i neka je

$$\mathcal{U}^{k-1} = span(u_1, \ldots, u_{k-1}),$$
$$\mathcal{Y}^k = span(y_1, \ldots, y_k).$$

Za $y \in \mathcal{Y}^k$ označimo $y' = \begin{bmatrix} y \\ 0 \end{bmatrix} \in \mathbb{R}^n$. Tada vrijedi $(y')^T (y') = y^T y$ i

$$(y')^T A (y') = [y^T \ 0] \begin{bmatrix} B & c \\ c^T & d \end{bmatrix} \begin{bmatrix} y \\ 0 \end{bmatrix} = [y^T \ 0] \begin{bmatrix} By \\ c^T y \end{bmatrix} = y^T B y. \tag{1.5}$$

Neka je $\mathcal{X}^k = \{ \begin{bmatrix} y \\ 0 \end{bmatrix} \in \mathbb{R}^n : y \in \mathcal{Y}^k \}$. Budući da su $\mathcal{X}^k$ i $\mathcal{U}^{k-1}$ potprostori od $\mathbb{R}^n$, $n < \infty$, te

$$\dim \mathcal{U}^{k-1} = k - 1, \quad \dim \mathcal{X}^k = k,$$

postoji $v \in (\mathcal{U}^{k-1})^{\perp} \cap \mathcal{X}^k$, $\|v\|_2 = 1$. Tada vrijedi

$$\begin{aligned}
\lambda_k &= \max_{\substack{x^T x=1 \\ x \perp \mathcal{U}^{k-1}}} x^T A x && \text{lema 1.2} \\
&\geq v^T A v && \text{jer je } v \perp \mathcal{U}^{k-1} \\
&\geq \min_{\substack{x^T x=1 \\ x \in \mathcal{X}^k}} x^T A x && \text{jer je } v \in \mathcal{X}^k \\
&= \min_{\substack{x^T x=1 \\ x \in \mathcal{Y}^k}} x^T B x && \text{zbog (1.5)} \\
&= \mu_k && \text{lema 1.2.}
\end{aligned}$$

Dakle,

$$\lambda_k \geq \mu_k, \ \forall k \in \{1, \ldots, n-1\}. \tag{1.6}$$

Kako bi dokazali drugu nejednakost,

$$\mu_k \geq \lambda_{k+1}, \ \forall k \in \{1, \ldots, n-1\}, \tag{1.7}$$

uvedimo notaciju

$$\lambda_{-j} \equiv \lambda_{n+1-j}, \quad \mu_{-j} \equiv \mu_{n-j}.$$

Sada (1.7) možemo zapisati kao

$$\mu_{-j} = \mu_{n-j} \geq \lambda_{n+1-j} = \lambda_{-j}, \ \forall j \in \{1, \ldots, n-1\}. \tag{1.8}$$

Svojstvene vrijednosti matrice $-A$ su

$$(-\lambda_{-1}) \geq \cdots \geq (-\lambda_{-(n-1)}) \geq (-\lambda_1),$$

a svojstvene vrijednosti matrice $-B$

$$(-\mu_{-1}) \geq \cdots \geq (-\mu_{-(n-1)}),$$

pa primjenom tvrdnje (1.6) za matricu $-A$ dobivamo upravo

$$(-\lambda_{-k}) \geq (-\mu_{-k}), \ \forall k \in \{1, \ldots, n-1\},$$

što je ekvivalentno (1.8) tj. (1.7). □

Teorem 1.4 (Courant - Fischer, [2, Teorem 8.1.2]). *Neka je $A \in \mathbb{R}^{n\times n}$ simetrična matrica i neka je s (1.1) dana njena spektralna dekompozicija. Tada za $k = 1, \ldots, n$ vrijedi*

$$\lambda_k = \min_{\mathcal{S}_{k-1}} \max_{\substack{x^T x=1 \\ x\perp\mathcal{S}_{k-1}}} x^T A x,$$

gdje $\mathcal{S}_{k-1}$ označava $(k-1)$ - dimenzionalan potprostor od $\mathbb{R}^n$.

Dokaz. Neka je $k \in \{1, \ldots, n\}$. Iz leme 1.2 slijedi

$$\lambda_k = \max_{\substack{x^T x=1 \\ x\perp u_1,\ldots,u_{k-1}}} x^T A x \geq \min_{\mathcal{S}_{k-1}} \max_{\substack{x^T x=1 \\ x\perp\mathcal{S}_{k-1}}} x^T A x.$$

Još treba dokazati obratnu nejednakost. Uzmimo proizvoljan $\mathcal{S}_{k-1} \leq \mathbb{R}^n$, $\dim \mathcal{S}_{k-1} = k-1$. Vrijedi $\dim \mathcal{S}_{k-1}^{\perp} = n-k+1$, $\dim[\{u_{k+1}, \ldots, u_n\}]^{\perp} = k$, pa je $\mathcal{S}_{k-1}^{\perp} \cap [\{u_{k+1}, \ldots, u_n\}]^{\perp} \neq \emptyset$. Uzmimo $y \in \mathcal{S}_{k-1}^{\perp} \cap [\{u_{k+1}, \ldots, u_n\}]^{\perp}$, $y^T y = 1$. Tada je

$$\begin{aligned}
\max_{\substack{x^T x=1 \\ x\perp\mathcal{S}_{k-1}}} x^T A x &\geq y^T A y && \text{jer je } y \perp \mathcal{S}_{k-1},\ y^T y = 1 \\
&\geq \min_{\substack{x^T x=1 \\ x\perp u_{k+1},\ldots,u_n}} x^T A x && \text{jer je } y \perp u_{k+1}, \ldots, u_n,\ y^T y = 1 \\
&= \lambda_k && \text{lema 1.2.}
\end{aligned}$$

Budući da je $\mathcal{S}_{k-1}$ bio proizvoljan $(k-1)$ - dimenzionalan podskup od $\mathbb{R}^n$, vrijedi

$$\min_{\mathcal{S}_{k-1}} \max_{\substack{x^T x=1 \\ x\perp\mathcal{S}_{k-1}}} x^T A x \geq \lambda_k.$$

Dakle,

$$\lambda_k = \min_{\mathcal{S}_{k-1}} \max_{\substack{x^T x=1 \\ x\perp\mathcal{S}_{k-1}}} x^T A x, \quad k = 1, \ldots, n.$$

□

Korolar 1.5 ([3]). *Neka su $A, E \in \mathbb{R}^{n\times n}$ simetrične matrice. Tada je*

$$|\lambda_k(A+E) - \lambda_k(A)| \leq \|E\|_2, \quad k = 1, \ldots, n.$$

Dokaz. Neka je $k \in \{1, \ldots, n\}$. Iz teorema 1.4 slijedi

$$\lambda_k(A+E) = \min_{\mathcal{S}_{k-1}} \max_{\substack{x^T x=1 \\ x \perp \mathcal{S}_{k-1}}} x^T(A+E)x = \min_{\mathcal{S}_{k-1}} \max_{\substack{x^T x=1 \\ x \perp \mathcal{S}_{k-1}}} (x^T Ax + x^T Ex).$$

Neka je (1.1) spektralna dekompozicija od A i neka je $q_i = Ue_i,\ i = 1, \ldots, k-1$. Tada je $\dim[\{q_1, \ldots, q_{k-1}\}] = k-1$, pa vrijedi

$$\lambda_k(A+E) \leq \max_{\substack{x^T x=1 \\ x \perp [\{q_1, \ldots, q_{k-1}\}]}} (x^T Ax + x^T Ex).$$

Neka je $i \in \{1, \ldots, k-1\}$. Iz $x \perp q_i$, slijedi $0 = q_i^T x = e_i^T (U^T x) \equiv e_i^T y$, tj. prvih $k-1$ komponenti vektora y su nula. Zato je

$$x^T Ax = x^T U \Lambda U^T x = y^T \Lambda y = \sum_{i=k}^{n} \lambda_i(A) y_i^2.$$

Budući da je

$$\sum_{i=k}^{n} \lambda_i(A) y_i^2 \leq \lambda_k(A), \quad x^T Ex \leq \lambda_1(E),$$

imamo

$$\lambda_k(A+E) \leq \lambda_k(A) + \lambda_1(E). \tag{1.9}$$

Kada prethodno dokazanu tvrdnju (1.9) primijenimo na $A = (A+E) - E$, dobivamo

$$\lambda_k(A) \leq \lambda_k(A+E) + (-\lambda_n(E)) \tag{1.10}$$

Iz (1.9) i (1.10) slijedi

$$|\lambda_k(A+E) - \lambda_k(A)| \leq \max\{|\lambda_1(E), \lambda_n(E)|\} = \|E\|_2.$$

□

Poglavlje 2

Spektralna dekompozicija streličastih matrica

U ovom poglavlju opisujemo algoritam za sprektralnu dekompoziciju streličastih matrica zvan *aheig* (ArrowHEad EIGenvalues/eigenvectors), preuzet iz [7] i [8]. Prvo definiramo streličastu matricu, te prezentiramo osnovnu verziju algoritma, *aheig_basic*, njegovu implementaciju u Matlabu i neke primjere. Zatim analiziramo točnost opisanog algoritma *aheig_basic*, dajemo ispravke u određenim 'lošim' slučajevima, te konačno dolazimo do algoritma *aheig*.

2.1 Algoritam *aheig_basic*

Za realnu simetričnu matricu $A \in \mathbb{R}^{n\times n}$ kažemo da je streličasta ako je oblika

$$A = \begin{bmatrix} D & z \\ z^T & \alpha \end{bmatrix}, \tag{2.1}$$

gdje je

$$D = diag(d_1, d_2, \ldots, d_{n-1})$$

dijagonalna matrica reda $n-1$,

$$z = [\zeta_1 \ \zeta_2 \ \cdots \ \zeta_{n-1}]^T$$

vektor, a α skalar.

Do kraja poglavlja 2 smatramo da je matrica $A \in \mathbb{R}^{n\times n}$ streličasta, oblika (2.1).

Bez smanjenja općenitosti možemo pretpostaviti da vrijedi:

1. $\zeta_i \neq 0,\ 1 \leq i \leq n-1,$

   Ako je $\zeta_i = 0$ za neki $i \in \{1, \ldots, n-1\}$, onda je (d_i, e_i) svojstveni par matrice A. Označimo s B matricu dobivenu brisanjem i-tog retka i i-tog stupca iz matrice A. Ako je $(\lambda, x) \in \mathbb{R} \times \mathbb{R}^{n-1}$ svojstveni par matrice B, onda vrijedi

$$\begin{bmatrix} d_1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \zeta_1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & d_{i-1} & 0 & 0 & \cdots & 0 & \zeta_{i-1} \\ 0 & \cdots & 0 & d_i & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & d_{i+1} & \cdots & 0 & \zeta_{i+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & d_{n-1} & \zeta_{n-1} \\ \zeta_1 & \cdots & \zeta_{i-1} & 0 & \zeta_{i+1} & \cdots & \zeta_{n-1} & \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ 0 \\ x_i \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ 0 \\ x_i \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{bmatrix}.$$

   Svi retci osim i-tog vrijede jer je (λ, x) svojstveni par matrice B, a i-ti redak jer je

$$d_i \cdot 0 = \lambda \cdot 0.$$

   Dakle, s $(\lambda, [x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_{n-1}]^T)$ je dan svojstveni par matrice A, pa je dovoljno naći spektralnu dekompoziciju matrice B. Postupak se može ponoviti za svaki $j \in \{1, \ldots, n\}$ za koji vrijedi $\zeta_j = 0$.

2. $d_i \neq d_j,\ i \neq j,\ 1 \leq i, j \leq n-1,$

   Neka je $d_i = d_j$ za neke $i \neq j$, $1 \leq i, j \leq n-1$. Cauchyjev teorem o preplitanju (teorem 1.3) daje

$$\lambda_1 \geq d_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq d_{n-1} \geq \lambda_n, \tag{2.2}$$

   pa je d_i svojstvena vrijednost od A, te se Givensovim rotacijama može poništiti element ζ_i. Neka je s

$$G_{(i,j)} = \begin{bmatrix} 1 & & & & & & & \\ & \cdots & & & & & & \\ & & \cos\psi & & & & \sin\psi & \\ & & & & 1 & & & \\ & & \vdots & & & \cdots & \vdots & \\ & & & & & & 1 & \\ & & -\sin\psi & & & & \cos\psi & \\ & & & & & & & \cdots \\ & & & & & & & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Matrica $G_{(i,j)}$ na dijagonali ima jedinice, osim na mjestima $g_{ii} = g_{jj} = \cos\varphi$, a izvan dijagonale nule, osim na $g_{ij} = \sin\varphi$, $g_{ji} = -\sin\varphi$. Kut ψ odaberemo tako da za matricu $G^T_{(i,j)}AG_{(i,i)} = [a'_{kl}]^n_{k,l=1}$ vrijedi

$$0 = a'_{in} = a'_{ni} = a_{in}\cos\psi - a_{jn}\sin\psi$$

tj.

$$\tan\psi = \frac{a_{in}}{a_{jn}} = \frac{\zeta_i}{\zeta_j}.$$

Dakle, možemo staviti

$$\sin\psi = \frac{a_{in}}{\sqrt{a_{in}^2 + a_{jn}^2}} = \frac{\zeta_i}{\sqrt{\zeta_i^2 + \zeta_j^2}}, \quad \cos\psi = \frac{a_{jn}}{\sqrt{a_{in}^2 + a_{jn}^2}} = \frac{\zeta_j}{\sqrt{\zeta_i^2 + \zeta_j^2}}.$$

Za ostale elemente u matrici $G^T_{(i,j)}AG_{(i,j)}$ tada vrijedi

$$\begin{aligned}
a'_{ii} &= a_{ii}\cos^2\psi - 2a_{ij}\sin\psi\cos\psi + a_{jj}\sin^2\psi \\
&= d_i\cos^2\psi + d_j\sin^2\psi \\
&= d_i, \quad \text{jer je } d_i = d_j,\ a_{ij} = 0,
\end{aligned}$$

$$\begin{aligned}
a'_{jj} &= a_{ii}\sin^2\psi + 2a_{ij}\sin\psi\cos\psi + a_{jj}\cos^2\psi \\
&= d_i\sin^2\psi + d_j\cos^2\psi \\
&= d_j = d_i, \quad \text{jer je } d_i = d_j,\ a_{ij} = 0,
\end{aligned}$$

$$\begin{aligned}
a'_{ir} = a'_{ri} &= a_{ir}\cos\psi - a_{jr}\sin\psi \\
&= 0,\ r \notin \{i, j, n\}, \quad \text{jer je } a_{ir} = a_{jr} = 0,
\end{aligned}$$

$$\begin{aligned}
a'_{jr} = a'_{rj} &= a_{rj}\cos\psi + a_{ri}\sin\psi \\
&= 0,\ r \notin \{i, j, n\}, \quad \text{jer je } a_{ri} = a_{rj} = 0,
\end{aligned}$$

$$\begin{aligned}
a'_{ij} = a'_{ji} &= a_{ij}(\cos^2\varphi - \sin^2\varphi) + (a_{ii} - a_{jj})\sin\varphi\cos\varphi, \\
&= (d_i - d_j)\sin\varphi\cos\varphi \\
&= 0, \quad \text{jer je } d_i = d_j,\ a_{ij} = 0,
\end{aligned}$$

$$a'_{rs} = a_{rs},\ r, s \notin \{i, j\},$$

$$\zeta'_j \equiv a'_{jn} = a'_{nj} = a_{nj}\cos\psi + a_{ni}\sin\psi = \sqrt{a_{nj}^2 + a_{ni}^2} = \sqrt{\zeta_j^2 + \zeta_i^2}.$$

Sve zajedno, dobivamo matricu

$$G_{(i,j)}^T A G_{(i,j)} = \begin{bmatrix} d_1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \zeta_1 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & d_i & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & d_i & \cdots & 0 & \zeta_j' \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & d_{n-1} & \zeta_{n-1} \\ \zeta_1 & \cdots & 0 & \cdots & \zeta_j' & \cdots & \zeta_{n-1} & \alpha \end{bmatrix},$$

pa možemo nastaviti kao u slučaju 1.

Matlab kod opisane rotacije je u Algoritmu 1.

Algoritam 1

```
function [A,Q] = Givens(A,i,j)
n = size(A,1);
t = sqrt(A(i,n)*A(i,n)+A(j,n)*A(j,n));
s = A(i,n)/t;
c = A(j,n)/t;
Q = eye(size(A)); Q(i,i) = c; Q(j,j) = c; Q(i,j) = s; Q(j,i) = -s

A(j,n) = t; A(n,j)=t;
A(i,n) = 0; A(n,i) = 0;
end
```

3. $d_1 > d_2 > \cdots > d_{n-1}$.
   Ako to ne vrijedi, moguće je naći permutaciju $p : \{1, \ldots, n-1\} \to \{1, \ldots, n-1\}$ takvu da vrijedi

$$d_{p(1)} > d_{p(2)} > \cdots > d_{p(n-1)}. \tag{2.3}$$

Neka je $P \in \mathbb{R}^{n \times n}$ matrica za koju vrijedi $Pe_i = e_{p(i)}$. Tada je

$$\begin{aligned} \text{za } i \in \{1, \ldots, n-1\}, \quad & P(p(i), i) = 1, \quad && P(j, i) = 0,\ j \neq p(i),\ j \in \{1, \ldots, n\}, \\ & P(n, n) = 1, && P(j, n) = 0,\ 1 \leq j \leq n-1. \end{aligned}$$

Dakle, matrica P ima u svakom retku i u svakom stupcu po jednu jedinicu, a ostalo nule i vrijedi

$$P^T P = \begin{bmatrix} P(p(1),:) \\ \vdots \\ P(p(n-1),:) \\ P(n,:) \end{bmatrix} = I_n,$$

tj. $P^{-1} = P^T$. Zato imamo

$$B = AP = \begin{bmatrix} A(:,p(1)) & \cdots & A(:,p(n-1)) & A(:,n) \end{bmatrix}$$

i

$$P^{-1}AP = \begin{bmatrix} B(p(1),:) \\ \vdots \\ B(p(n-1),:) \\ B(n,:) \end{bmatrix} = \begin{bmatrix} d_{p(1)} & \cdots & 0 & \zeta_{p(1)} \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & d_{p(n-1)} & \zeta_{p(n-1)} \\ \zeta_{p(1)} & \cdots & \zeta_{p(n-1)} & \zeta_n \end{bmatrix}.$$

Dakle, matrica $P^{-1}AP$ ima traženo svojstvo (2.3), a iz

$$P^{-1}APx = \lambda x \iff A(Px) = \lambda(Px)$$

tj.

$$P^{-1}AP = U\Lambda U^T \iff A = (PU)\Lambda(PU)^T.$$

se vidi da je dovoljno je naći spektralnu dekompoziciju matrice $P^{-1}AP$.

Kažemo da je A reducirana ako vrijedi 1 i 2. Ako vrijedi 3, kažemo da je uređena. U nastavku poglavlja smatramo da je A uređena i reducirana.

Teorem 2.1 ([10])**.** *Neka je A streličasta matrica oblika (2.1) koja je uređena i reducirana, i neka je s (1.1) dana njena spektralna dekompozicija. Neka je*

$$d_0 > \max\{d_1 + |\zeta_1|, \ldots, d_{n-1} + |\zeta_{n-1}|, \alpha + \sum_{i=1}^{n-1} |\zeta_i|\}, \tag{2.4}$$

$$d_n < \min\{d_1 - |\zeta_1|, \ldots, d_{n-1} - |\zeta_{n-1}|, \alpha - \sum_{i=1}^{n-1} |\zeta_i|\}, \tag{2.5}$$

i neka je $f_A : \mathbb{R} \to \mathbb{R}$ *funkcija definirana s*

$$f_A(x) = \alpha - x - \sum_{i=1}^{n-1} \frac{\zeta_i^2}{d_i - x} = \alpha - x - z^T(D - xI)^{-1}z, \ \forall x \in \mathbb{R}. \tag{2.6}$$

Tada vrijedi

$$d_0 > \lambda_1 > d_1 > \lambda_2 > \cdots > \lambda_{n-1} > d_{n-1} > \lambda_n > d_n \tag{2.7}$$

i

$$\lambda \text{ je svojstvena vrijednost matrice } A \text{ ako i samo ako vrijedi } f_A(\lambda) = 0. \tag{2.8}$$

Svojstveni vektori $u_1, \ldots, u_n$ *su dani s*

$$u_i = \frac{x_i}{\|x_i\|_2}, \quad x_i = \begin{bmatrix} (D - \lambda_i I)^{-1} z \\ -1 \end{bmatrix}, \quad i = 1, \ldots, n.$$

Dokaz. Prvo ćemo pokazati da za $\lambda \in \mathbb{R}$ koji je različit od d_i, $1 \leq i \leq n-1$, vrijedi (2.8). Neka je $\lambda \neq d_i, \ \forall i \in \{1, \ldots, n-1\}$. Tada je λ svojstvena vrijednost od A ako i samo ako je $\det(A - \lambda I) = 0$ tj.

$$\begin{vmatrix} d_1 - \lambda & 0 & \ldots & 0 & \zeta_1 \\ 0 & d_2 - \lambda & & 0 & \zeta_2 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & d_{n-1} - \lambda & \zeta_{n-1} \\ \zeta_1 & \zeta_2 & \ldots & \zeta_{n-1} & \alpha - \lambda \end{vmatrix} = 0.$$

Ako za $i = 1, \ldots, n-1$ redom zadnjem retku dodajemo i-ti redak pomnožen sa $-\zeta_i/(d_i - \lambda)$ (to možemo jer je λ različit od svih d_i), determinanta se neće promijeniti i dobivamo jednadžbu

$$\begin{vmatrix} d_1 - \lambda & 0 & \ldots & 0 & \zeta_1 \\ 0 & d_2 - \lambda & & 0 & \zeta_2 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & d_{n-1} - \lambda & \zeta_{n-1} \\ 0 & 0 & \ldots & 0 & f_A(\lambda) \end{vmatrix} = 0.$$

Determinanta trokutaste matrice je produkt elemenata na dijagonali, pa, budući da je λ različit od svih d_i, mora vrijediti $f_A(\lambda) = 0$. Obratno, $f_A(\lambda) = 0$ povlači $\det(A - \lambda I) = 0$, pa je λ svojstvena vrijednost od A.

Sada uzmimo $i \in \{1, \ldots, n-1\}$. Prema pretpostavci vrijedi $d_i > d_{i+1}$. Za $\lambda \in \mathbb{R}$ koji je dovoljno blizu d_i, a ipak od njega manji vrijedi (jer je $\zeta_i \neq 0$)

$$f_A(\lambda) \approx -\frac{\zeta_i^2}{d_i - \lambda} < 0.$$

Za λ koji je dovoljno blizu d_{i+1}, a ipak od njega veći vrijedi

$$f_A(\lambda) \approx -\frac{\zeta_{i+1}^2}{d_{i+1} - \lambda} > 0.$$

Slijedi da f mijenja predznak na $\langle d_{i+1}, d_i \rangle$. Iz prve tvrdnje se vidi da se to dogodi u svojstvenoj vrijednosti matrice A. S obzirom da otvorenih intervala $\langle d_{i+1}, d_i \rangle$ ima ukupno $n-2$ i u svakom je jedna svojstvena vrijednost, preostaje pokazati da postoji svojstvena vrijednost veća od d_1 i svojstvena vrijednost manja od d_n. Ako je λ dovoljno blizu d_1, ali je od nje veća, vrijedi

$$f_A(\lambda) \approx -\frac{\zeta_1^2}{d_1 - \lambda} > 0.$$

S druge strane, za dovoljno velike λ imamo $f_A(\lambda) \approx -\lambda < 0$, pa f mijenja predznak na $\langle d_1, \infty \rangle$. Analogno, ako je λ dovoljno blizu d_n, a od nje manja, vrijedi

$$f_A(\lambda) \approx -\frac{\zeta_n^2}{d_n - \lambda} < 0,$$

a za dovoljno male λ (tj. za $\lambda < 0$, koji je dovoljno velik po modulu) je $f_A(\lambda) \approx -\lambda > 0$, pa f_A mijenja predznak na $\langle -\infty, d_n \rangle$.
Iz teorema 1.1 slijedi

$$\begin{aligned}
\sigma(A) &\subseteq \bigcup_{j=1}^{n} [a_{jj} - \sum_{\substack{k=1 \\ k \neq j}}^{n} |a_{jk}|,\ a_{jj} + \sum_{\substack{k=1 \\ k \neq j}}^{n} |a_{jk}|] \\
&= \bigcup_{j=1}^{n-1} [d_{jj} - |\zeta_j|,\ d_{jj} + |\zeta_j|] \cup [\alpha - \sum_{i=1}^{n-1} |\zeta_i|,\ \alpha + \sum_{i=1}^{n-1} |\zeta_i|] \\
&\subseteq [\min\{d_1 - |\zeta_1|, \ldots, d_{n-1} - |\zeta_{n-1}|,\ \alpha - \sum_{i=1}^{n-1} |\zeta_i|\}, \\
&\qquad \max\{d_1 + |\zeta_1|, \ldots, d_{n-1} + |\zeta_{n-1}|,\ \alpha + \sum_{i=1}^{n-1} |\zeta_i|\}] \\
&\subset \langle d_n, d_0 \rangle.
\end{aligned}$$

Time je pokazano da vrijedi

$$d_0 > \lambda_1 > d_1 > \lambda_2 > \cdots > \lambda_{n-1} > d_{n-1} > \lambda_n > d_n$$

i pritom su svojstvene vrijednosti matrice A nultočke funkcije f.

Neka je $i \in \{1, \ldots, n\}$. Tada vrijedi

$$0 = f_A(\lambda_i) = \alpha - \lambda_i - z^T(D - \lambda_i I)^{-1} z$$

i

$$\begin{aligned}(D(D - \lambda_i I)^{-1} - I) &= \operatorname{diag}\left(\frac{d_1}{d_1 - \lambda_i} - 1, \ldots, \frac{d_{n-1}}{d_{n-1} - \lambda_i} - 1\right) \\ &= \operatorname{diag}\left(\frac{\lambda_i}{d_1 - \lambda_i}, \ldots, \frac{\lambda_i}{d_{n-1} - \lambda_i}\right) \\ &= \lambda_i (D - \lambda_i I)^{-1},\end{aligned}$$

pa je

$$\begin{bmatrix} D & z \\ z^T & \alpha \end{bmatrix} \begin{bmatrix} (D - \lambda_i I)^{-1} z \\ -1 \end{bmatrix} = \begin{bmatrix} (D(D - \lambda_i I)^{-1} - I)z \\ z^T (D - \lambda_i I)^{-1} z - \alpha \end{bmatrix} = \lambda_i \begin{bmatrix} (D - \lambda_i I)^{-1} z \\ -1 \end{bmatrix}.$$

□

Vrijednosti d_i, $i \in \{1, \ldots, n-1\}$ zovemo polovi.

Sada opisujemo ideju algoritma *aheig_basic*. Neka je λ svojstvena vrijednost matrice A,

$$x = \begin{bmatrix} (D - \lambda I)^{-1} z \\ -1 \end{bmatrix} \tag{2.9}$$

odgovarajući svojstveni vektor i $u = x/\|x\|_2$ odgovarajući normirani svojstveni vektor. Neka je $i \in \{1, \ldots, n-1\}$ indeks za koji je pol d_i najbliži svojstvenoj vrijednosti λ. Iz (2.7) slijedi $\lambda = \lambda_i$ ili $\lambda = \lambda_{i+1}$. Tada definiramo matricu A_i s

$$A_i = A - d_i I = \begin{bmatrix} D_1 & 0 & 0 & z_1 \\ 0 & 0 & 0 & \zeta_i \\ 0 & 0 & D_2 & z_2 \\ z_1^T & \zeta_i & z_2^T & a \end{bmatrix}, \tag{2.10}$$

gdje je

$$\begin{aligned} D_1 &= \operatorname{diag}(d_1 - d_i, \ldots, d_{i-1} - d_i) \quad \text{pozitivno definitna matrica,} \\ D_2 &= \operatorname{diag}(d_{i+1} - d_i, \ldots, d_{n-1} - d_i) \quad \text{negativno definitna matrica,} \\ z_1 &= [\zeta_1 \ \cdots \ \zeta_{i-1}]^T, \\ z_2 &= [\zeta_{i+1} \ \cdots \ \zeta_{n-1}]^T, \\ a &= \alpha - d_i. \end{aligned} \tag{2.11}$$

Lema 2.2. *Inverz matrice A_i je dan s*

$$A_i^{-1} = \begin{bmatrix} D_1^{-1} & w_1 & 0 & 0 \\ w_1^T & b & w_2^T & 1/\zeta_i \\ 0 & w_2 & D_2^{-1} & 0 \\ 0 & 1/\zeta_i & 0 & 0 \end{bmatrix}, \tag{2.12}$$

gdje je

$$\begin{aligned} w_1 &= -D_1^{-1} z_1 \frac{1}{\zeta_i}, \\ w_2 &= -D_2^{-1} z_2 \frac{1}{\zeta_i}, \\ b &= \frac{1}{\zeta_i^2}(-a + z_1^T D_1^{-1} z_1 + z_2^T D_2^{-1} z_2). \end{aligned} \tag{2.13}$$

Dokaz. Računamo

$$\begin{bmatrix} D_1^{-1} & w_1 & 0 & 0 \\ w_1^T & b & w_2^T & 1/\zeta_i \\ 0 & w_2 & D_2^{-1} & 0 \\ 0 & 1/\zeta_i & 0 & 0 \end{bmatrix} \begin{bmatrix} D_1 & 0 & 0 & z_1 \\ 0 & 0 & 0 & \zeta_i \\ 0 & 0 & D_2 & z_2 \\ z_1^T & \zeta_i & z_2^T & a \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ B_{31} & B_{32} & B_{33} & B_{34} \\ B_{41} & B_{42} & B_{43} & B_{44} \end{bmatrix}.$$

Odmah se vidi da je

$$B_{11} = D_1^{-1} D_1 = I_{i-1}, \quad B_{33} = D_2^{-1} D_2 = I_{n-i-1}, \quad B_{22} = B_{44} = \frac{1}{\zeta_i}\zeta_i = 1,$$

te

$$\begin{aligned} &B_{12} = 0_{(i-1)\times 1}, \quad B_{13} = 0_{(i-1)\times(n-i-1)}, \quad B_{31} = 0_{(n-i-1)\times(i-1)}, \quad B_{32} = 0_{(n-i-1)\times 1}, \\ &B_{41} = B_{42} = B_{43} = 0. \end{aligned}$$

Zbog $w_1 = -D_1^{-1} z_1 \frac{1}{\zeta_i}$ i činjenice da je D_1 dijagonalna matrica, pa vrijedi $D_1 = D_1^T$, je

$$\begin{aligned} B_{14} &= D_1^{-1} z_1 + w_1 \zeta_i = D_1^{-1} z_1 + (-D_1^{-1} z_1 \frac{1}{\zeta_i})\zeta_i = 0, \\ B_{21} &= w_1^T D_1 + \frac{1}{\zeta_i} z_1^T = (-\frac{1}{\zeta_i} z_1^T D_1^{-T}) D_1 + \frac{1}{\zeta_i} z_1^T = 0. \end{aligned}$$

Zbog $w_2 = -D_2^{-1} z_2 \frac{1}{\zeta_i}$ i činjenice da je D_2 dijagonalna matrica, pa vrijedi $D_2 = D_2^T$, je

$$\begin{aligned} B_{34} &= w_2 \zeta_i + D_2^{-1} z_2 = (-D_2^{-1} z_2 \frac{1}{\zeta_i})\zeta_i + D_2^{-1} z_2 = 0, \\ B_{23} &= w_2^T D_2 + \frac{1}{\zeta_i} z_2^T = (-\frac{1}{\zeta_i} z_2^T D_2^{-T}) D_2 + \frac{1}{\zeta_i} z_2^T = 0. \end{aligned}$$

Budući da vrijedi $b = \frac{1}{\zeta_i^2}(-a + z_1^T D_1^{-1} z_1 + z_2^T D_2^{-1} z_2)$ i $w_i^T z_i = z_i^T w_i$, $i = 1, 2$, imamo

$$\begin{aligned} B_{24} &= w_1^T z_1 + b\zeta_i + w_2^T z_2 + \frac{1}{\zeta_i} a \\ &= w_1^T z_1 + \frac{1}{\zeta_i}(-a + z_1^T D_1^{-1} z_1 + z_2^T D_2^{-1} z_2) + w_2^T z_2 + \frac{1}{\zeta_i} a = 0. \end{aligned}$$

□

Lema 2.3. *Sljedeće tvrdnje su međusobno ekvivalentne:*

(*i*) (λ, x) *je svojstveni par matrice* A,

(*ii*) $(\mu, x) = (\lambda - d_i, x)$ *je svojstveni par matrice* A_i,

(*iii*) $(\nu, x) = (1/\mu, x)$ *je svojstveni par matrice* A_i^{-1}.

Dokaz. (*i*) $\Rightarrow$ (*ii*) Ako je λ svojstvena vrijednost od A, za pripadni svojstveni vektor x vrijedi

$$A_i x = (A - d_i I)x = (\lambda - d_i)x = \mu x,$$

tj. μ je svojstvena vrijednost matrice A_i uz isti svojstveni vektor.
(*ii*) $\Rightarrow$ (*i*) Vrijedi i obrat: ako je $A_i y = \mu y$, onda je $\lambda = \mu + d_i$ svojstvena vrijednost matrice A uz isti svojstveni vektor jer vrijedi

$$\mu y = A_i y = (A - d_i I)y = Ay - d_i y \Rightarrow Ay = (\mu + d_i)y = \lambda y.$$

(*ii*) $\iff$ (*iii*) Skalar μ je svojstvena vrijednost matrice A_i ako i samo ako je $\nu = 1/\mu$ svojstvena vrijednost od A_i^{-1}, za isti svojstveni vektor jer je

$$A_i y = \mu y \iff \nu y = A_i^{-1} y,\ y \neq 0.$$

□

Iz prethodnih definicija i leme 2.3 odmah slijedi da vektor x definiran u (2.9) možemo zapisati i kao

$$x = \begin{bmatrix} \begin{bmatrix} D_1 - \mu I_{i-1} & 0 & 0 \\ 0 & -\mu & 0 \\ 0 & 0 & D_2 - \mu I_{n-i-1} \end{bmatrix}^{-1} \begin{bmatrix} z_1 \\ \zeta_i \\ z_2 \end{bmatrix} \\ -1 \end{bmatrix} = \begin{bmatrix} (D_1 - \mu I)^{-1} z_1 \\ -\frac{\zeta_i}{\mu} \\ (D_2 - \mu I)^{-1} z_2 \\ -1 \end{bmatrix}. \tag{2.14}$$

Sada ćemo prezentirati implementaciju algoritma *aheig_basic* u Matlabu, detaljno opisati svrhu pojedinih linija koda, te pritom objasniti algoritam. Budući da se svaki svojstveni par može naći potpuno nezavisno od ostalih, opisivat ćemo verziju algoritma koja nalazi samo jedan svojstveni par (λ_k, u_k).

Matlab funkcija *aheig_basic* (Algoritam 2):

ULAZ: $D \in \mathbb{R}^{n-1}$, $z \in \mathbb{R}^{n-1}$, $\alpha \in \mathbb{R}$, indeks tražene svojstvene vrijednosti k.

2 Tražimo svojstveni par (λ_k, u_k) matrice

$$A = \begin{bmatrix} diag(D) & z \\ z^T & \alpha \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

3-23 Odredimo indeks $i \in \{1, \ldots, n-1\}$, za koji je d_i najbliži pol traženoj svojstvenoj vrijednosti λ_k. Odmah zapamtimo pomak d_i u varijablu $shift$, te s koje strane strane pola d_i se nalazi vrijednost λ_k ('R' ako je $d_i < \lambda_k$, a 'L' ako je $\lambda_k < d_i$). Pritom koristimo (2.7):

$$\lambda_1 > d_1 > \lambda_2 > \cdots > \lambda_{n-1} > d_{n-1} > \lambda_n.$$

3-6 Ako je $k = 1$ tražimo najveću svojstvenu vrijednost λ_1, pa je najbliži pol d_1.

7-10 Za $k = n$, tražimo najmanju svojstvenu vrijednost λ_n kojoj je najbliži pol d_{n-1}.

11-22 Ukoliko je $2 \leq k \leq n-1$, najbliži pol može biti d_k ili d_{k-1}. Koristimo funkciju f_A i činjenice iz dokaza teorema 2.1. Funkcija f_A je neprekidna na intervalu $\langle d_k, d_{k-1}\rangle$, a za brojeve $x \in \mathbb{R}$ koji su dovoljno blizu polu d_k i od njega veći je $f_A(x) > 0$. Ako je x dovoljno blizu d_{k-1}, a od njega manji, onda je $f_A(x) < 0$. Također znamo da f_A mijenja predznak u λ_k (ilustracija za konkretnu matricu je na Slici 2.1). Zato izračunamo vrijednost

$$f_A\left(\frac{d_{k-1} + d_k}{2}\right).$$

Ako je ona pozitivna, zaključujemo da se promjena predznaka dogodi na intervalu $\langle (d_{k-1}+d_k)/2, d_{k-1}\rangle$, pa je najbliži pol d_{k-1}. Ako je negativna, onda se predznak promijeni na intervalu $\langle d_k, (d_{k-1}+d_k)/2\rangle$. Kada bi bilo $f_A\Big((d_{k-1}+d_k)/2\Big) = 0$, onda bi tražena svojstvena vrijednost bila $\lambda_k = (d_{k-1} + d_k)/2$.

24 Pozivamo Matlab funkciju *invA* (Algoritam 3) koja računa matricu

$$A_i^{-1} = (A - d_i)^{-1}$$

prema (2.11), (2.12) i (2.13).

25 Tražimo svojstvenu vrijednost λ_k matrice A i koristimo tvrdnju leme 2.3:

$$\lambda_k \in \sigma(A) \iff \mu_k = \lambda_k - d_i \in \sigma(A_i) \iff \nu_k = \frac{1}{\mu_k} \in \sigma(A_i^{-1}). \tag{2.15}$$

Polu d_i su najbliže svojstvene vrijednosti λ_i i λ_{i+1}. Iz (2.15) slijedi da su tada $\mu_i = \lambda_i - d_i$ i $\mu_{i+1} = \lambda_{i+1} - d_i$ svojstvene vrijednosti matrice A_i koje su najbliže nuli, te da je $\mu_i > 0$, a $\mu_{i+1} < 0$. Iz (2.15) se također se vidi da je $\nu_i = 1/\mu_i > 0$ najveća svojstvena vrijednost, a $\nu_{i+1} = 1/\mu_{i+1} < 0$ najmanja svojstvena vrijednost matrice A_i^{-1}.

Budući da je d_i pronađen kao pol koji je najbliži svojstvenoj vrijednosti λ_k, vrijedi $k \in \{i, i+1\}$. Ako je $\lambda_k < d_i$, onda je $\mu_k = \mu_{i+1} < 0$, pa tražimo najmanju svojstvenu vrijednost matrice A_i^{-1}, a ako je $d_i < \lambda_k$, onda je $\mu_k = \mu_i > 0$ i tražimo najveću svojstvenu vrijednost od A_i^{-1}. Pritom svojstvena vrijednost ν_k može biti po modulu najveća ili druga najveća svojstvena vrijednost matrice A_i^{-1}.

Svojstvenu vrijednost ν_k računamo rješavanjem jednadžbe

$$0 = f(x) \equiv \alpha - x - \sum_{i=1}^{n-1} \frac{\zeta_i^2}{d_i - x} = \alpha - x - z^T(D - xI)^{-1}z, \ x \in \mathbb{R}. \tag{2.16}$$

Pritom smo koristili činjenicu da se zamjenom i-tog i n-tog stupca, te i-tog i n-tog retka matrica A_i^{-1} može svesti na streličastu matricu. Neka je $P_{(i,n)} \in \mathbb{R}^{n\times n}$ matrica koja u svakom retku i svakom stupcu ima po jednu jedinicu, a ostalo nule i to

$$P_{(i,n)}(i,n) = P_{(i,n)}(n,i) = 1, P_{(i,n)}(j,j) = 1, j \notin \{i,n\}.$$

Tada je matrica

$$B = P_{(i,n)}^T A_i^{-1} P_{(i,n)},$$

streličasta, te vrijedi

$$By = \beta y \iff A_i^{-1}(P_{(i,n)}y) = \beta(P_{(i,n)}y).$$

Iz $\sigma(A_i^{-1}) = \sigma(B)$ i tvrdnje teorema 2.1 zaključujemo da su nultočke funkcije $f = f_B$ upravo svojstvene vrijednosti matrice A_i^{-1}.

Nultočku ν_k funkcije f računamo metodom bisekcije za matricu B pozivom Matlab funkcije *bisect* (Algoritam 4).

26 Računamo $\mu_k = 1/\nu_k$.

27 Svojstveni vektor x_k pridružen svojstvenoj vrijednosti μ_k računamo prema (2.14). Za to pozivamo Matlab funkciju *vect* (Algoritam 5) za matricu A_i i svojstvenu vrijednost μ_k. Nakon što izračuna x_k, funkcija vrati normirani svojstveni vektor $u_k = x_k/\|x_k\|_2$.

28 Računamo svojstvenu vrijednost λ_k matrice A kao $\lambda_k = \mu_k + d_i$.

IZLAZ: svojstvena vrijednost λ_k i pripadni normirani svojstveni vektor u_k.

Algoritam 2

```
function [lambda,u] = aheig_basic(D,z,alpha,k)
n = length(D)+1;
if k == 1
    shift = D(1);
    i = 1;
    side = 'R';
elseif k == n
    shift = D(n-1);
    i = n-1;
    side = 'L';
else
    middle = (D(k-1)+D(k))/2;
    Fmiddle = alpha - middle - sum(z.^2./(D-middle));
    if Fmiddle < 0
        shift = D(k);
        i = k;
        side = 'R';
    else
        shift = D(k-1);
        i = k-1;
        side = 'L';
    end
end
[invD1,invD2,w1,w2,wKsi,b] = invA(D,z,alpha,i);
lambda = bisect([invD1;0;invD2],[w1;wKsi;w2],b,side);
lambda = 1/lambda;
u = vect(D-shift,z,lambda);
lambda = lambda + shift;
end
```

Algoritam 3

```
function [invD1,invD2,w1,w2,wKsi,b] = invA(D,z,alpha,i)
n = length(D)+1;
a = alpha-D(i);
D = D-D(i);
invD1 = 1./D(1:i-1);
invD2 = 1./D(i+1:n-1);
wKsi = 1/z(i);
w1 = -z(1:i-1).*invD1*wKsi;
w2 = -z(i+1:n-1).*invD2*wKsi;
b = (-a+sum(z(1:i-1).^2.*invD1)+sum(z(i+1:n-1).^2.*invD2))*wKsi*wKsi;
end
```

Algoritam 4

```
function lambda = bisect(D,z,alpha,side)
if side == 'L'
    left = min(min(D-abs(z)),alpha-norm(z,1));
    right = min(D);
else
    right = max(max(D+abs(z)),alpha+norm(z,1));
    left = max(D);
end
middle = (left+right)/2;
while(right-left)>2*eps*max(abs(left),abs(right))
    Fmiddle = alpha - middle - sum(z.^2./(D-middle));
    if Fmiddle>0
        left = middle;
    else
        right = middle;
    end
    middle =(left+right)/2;
end
lambda = middle;
```

Algoritam 5

```
function v = vect(D,z,lambda)
v = [z./(D-lambda);-1];
v = v/norm(v,2);
end
```

Matlab funkcija *invA* (Algoritam 3):

ULAZ: dijagonala D, vektor z, skalar α, indeks najbližeg pola i.

3-10 Za danu matricu

$$A = \begin{bmatrix} diag(D) & z \\ z^T & \alpha \end{bmatrix}$$

i indeks $i \in \{1, \ldots, n-1\}$ računamo matricu A_i^{-1} prema (2.11), (2.12) i (2.13).

IZLAZ: dijagonala matrica D_1^{-1}, D_2^{-1}, vektori w_1, $w2$, skalari $w_\xi = 1/\zeta_i$, b.

Matlab funkcija *bisect* (Algoritam 4):

ULAZ: dijagonala D, vektor z, skalar α, varijabla $side \in \{'L', 'R'\}$.

2-8 Prvo odredimo interval na kojem ćemo tražiti nultočku funkcije f_C za matricu

$$C = \begin{bmatrix} diag(D) & z \\ z^T & \alpha \end{bmatrix}.$$

2-4 Ako je $side = 'L'$, tražimo najmanju svojstvenu vrijednost matrice C. Znamo da je ona manja od svih polova, pa za desni rub intervala uzimamo najmanji pol koji u Matlabu dobijemo s $min(D)$. Zbog (2.7) za lijevi rub uzimamo vrijednost d_n definiranu u (2.5).

5-7 Ako je $side = 'R'$, tražimo najveću svojstvenu vrijednost matrice C. Budući da je ona veća od svih polova, za lijevi rub intervala uzimamo najveći pol koji u Matlabu dobijemo s $max(D)$. Zbog (2.7) za desni rub uzimamo vrijednost d_0 definiranu u (2.4).

9-19 Metodom bisekcije na intervalu $\langle left, right\rangle$ tražimo nultočku funkcije f_C. U svakom koraku izračunamo

$$f_C(m), \quad m \equiv m = \frac{left + right}{2}.$$

Ako je vrijednost $f_C(m)$ pozitivna, znamo da se nultočka nalazi na intervalu $\langle m, right\rangle$, pa stavimo $left = m$. Ako je $f_C(m) < 0$, onda se nultočka nalazi u $\langle left, m\rangle$, pa stavimo $right = m$. Ako je $f_C(m) = 0$, našli smo nultočku m. Postupak ponavljamo dok slučajno ne nađemo nultočku ili dok interval $\langle left, right\rangle$ ne postane dovoljno mali. Tada sredinu tog intervala smatramo dobrom aproksimacijom tražene nultočke, pa stavimo $lambda = m$.

IZLAZ: skalar λ.
Ako je *side* = 'L', λ je najmanja svojstvena vrijednost matrice

$$C = \begin{bmatrix} diag(D) & z \\ z^T & \alpha \end{bmatrix},$$

inače je vraćena vrijednost λ najveća svojstvena vrijednost matrice C.

Matlab funkcija *vect* (Algoritam 5):

ULAZ: dijagonala D, vektor z, skalar λ.

2 Računamo svojstveni vektor x pridružen svojstvenoj vrijednosti λ matrice

$$\begin{bmatrix} diag(D) & z \\ z^T & \alpha \end{bmatrix}$$

prema (2.9):

$$x = \begin{bmatrix} (diag(D) - \lambda I)^{-1} z \\ -1 \end{bmatrix}.$$

3 Računamo $v = x/\|x\|_2$.

IZLAZ: vektor v.

Primjer 2.4. *Neka je zadana matrica*

$$A = \begin{bmatrix} 8 & 0 & 0 & 3 \\ 0 & 4 & 0 & 2 \\ 0 & 0 & 3 & 1 \\ 3 & 2 & 1 & 5 \end{bmatrix},$$

tj.

$$D = [8\ 4\ 3], \quad z = [3\ 2\ 1]^T, \quad \alpha = 5.$$

Svojstvene vrijednosti matrice A su (prema Wolfram Mathematici):

$$\begin{aligned} \lambda_1 &= 10.094421988827185, \\ \lambda_2 &= 5.2748074542539303, \\ \lambda_3 &= 3.1977211149230787, \\ \lambda_4 &= 1.4330494419958059. \end{aligned}$$

Matlab funkcija eig *daje*

$$\begin{aligned}\lambda_1 &= 10.094421988827182,\\ \lambda_2 &= 5.2748074542539305,\\ \lambda_3 &= 3.1977211149230778,\\ \lambda_4 &= 1.4330494419958042.\end{aligned}$$

Koristeći algoritam aheig_basic *svojstvene vrijednosti računamo na sljedeći način:*

- *Neka je* $k = 1$. *Računamo najveću svojstvenu vrijednost* λ_1, *pa je najbliži pol* $d_1 = 8$, *te prema (2.10) formiramo matricu*

$$A_1 = A - d_1 I = \begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & -4 & 0 & 2 \\ 0 & 0 & -5 & 1 \\ 3 & 2 & 1 & -3 \end{bmatrix}.$$

  Kako je $\mu_1 = \lambda_1 - d_1 > 0$, $\nu_1 = 1/\mu_1 > 0$ *je najveća (iako ne nužno po modulu najveća) svojstvena vrijednost matrice* A_1^{-1}, *koju računamo prema (2.12), (2.13). Dakle, za matricu* A_1^{-1}, *koja izračunata u Matlabu iznosi*

$$A_1^{-1} = \begin{bmatrix} 0.2 & 1.666666666666667 & 0.06666666666666667 & 0.3333333333333333 \\ 1.666666666666667 & -0.25 & 0 & 0 \\ 0.06666666666666667 & 0 & -0.2 & 0 \\ 0.3333333333333333 & 0 & 0 & 0 \end{bmatrix},$$

  metodom bisekcije računamo najveću svojstvenu vrijednost i ona iznosi

$$\nu_1 = 0.4774586999824093.$$

  Konačno dobivamo

$$\lambda_1 = \frac{1}{\nu_1} + d_1 = 10.094421988827186.$$

- *Neka je* $k = 4$. *Računamo najmanju svojstvenu vrijednost* λ_4, *pa je najbliži pol* $d_3 = 3$, *te prema (2.10) formiramo matricu*

$$A_3 = A - d_3 I = \begin{bmatrix} 5 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 3 & 2 & 1 & 2 \end{bmatrix}.$$

Kako je $\mu_4 = \lambda_4 - d_3 < 0$, $\nu_4 = 1/\mu_4 < 0$ *je najmanja (iako po modulu najveća ili druga najveća) svojstvena vrijednost matrice* A_3^{-1}, *koju računamo prema (2.12), (2.13). Dakle, za matricu*

$$A_3^{-1} = \begin{bmatrix} 0.2 & 0 & -0.6 & 0 \\ 0 & 1 & -2 & 0 \\ -0.6 & -2 & 3.8 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

metodom bisekcije računamo najmanju svojstvenu vrijednost i ona iznosi

$$\nu_4 = -0.6381822291021664.$$

Dobivamo

$$\lambda_4 = \frac{1}{\nu_4} + d_3 = 1.4330494419958060.$$

- *Neka je* $k = 2$. *Prvo tražimo najbliži pol, koristeći funkciju* f_A *(vidjeti sliku 2.1). Vrijedi*

$$f_A(\frac{d_1 + d_2}{2}) < 0,$$

pa zaključujemo da se nultočka nalazi u intervalu $\langle d_2, (d_1 + d_2)/2\rangle$ *tj. da je najbliži pol* $d_2 = 4$. *Prema (2.10) formiramo matricu*

$$A_2 = A - d_2 I = \begin{bmatrix} 4 & 0 & 0 & 3 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & -1 & 1 \\ 3 & 2 & 1 & 1 \end{bmatrix}.$$

Prema (2.12) i (2.13) računamo matricu

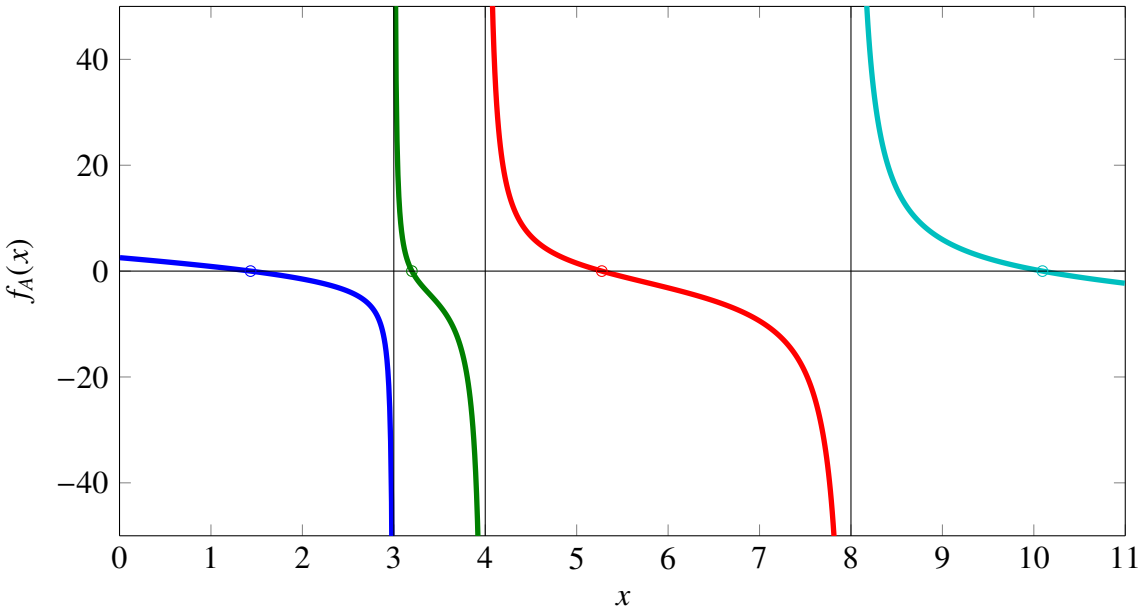
$$A_2^{-1} = \begin{bmatrix} 0.25 & -0.375 & 0 & 0 \\ -0.375 & 0.625 & 0.5 & 0.5 \\ 0 & 0.5 & -1 & 0 \\ 0 & 0.5 & 0 & 0 \end{bmatrix}.$$

Vrijedi $\mu_2 = \lambda_2 - d_2 > 0$, *pa metodom bisekcije računamo najveću svojstvenu vrijednost matrice* A_2^{-1}

$$\nu_2 = 0.7844321875143421.$$

Tada je

$$\lambda_2 = \frac{1}{\nu_2} + d_2 = 5.2748074542539305.$$



Slika 2.1: $f_A(x) = \alpha - x - \sum_{j=1}^{n-1} \frac{\zeta_j^2}{d_j - x},\ x \in \mathbb{R}$.

- *Neka je $k = 3$. Tražimo najbliži pol pomoću funkcije f_A (Slika 2.1). Izračunamo*

$$f_A(\frac{d_2 + d_3}{2}) < 0,$$

pa zaključujemo da je najbliži pol $d_3 = 3$, te prema (2.10) formiramo matricu

$$A_3 = A - d_3 I = \begin{bmatrix} 5 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 3 & 2 & 1 & 2 \end{bmatrix}.$$

Prema (2.12) i (2.13) računamo matricu

$$A_3^{-1} = \begin{bmatrix} 0.2 & 0 & -0.6 & 0 \\ 0 & 1 & -2 & 0 \\ -0.6 & -2 & 3.8 & 1.0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Vrijedi $\mu_3 = \lambda_3 - d_3 > 0$, pa računamo najveću svojstvenu vrijednost matrice A_3^{-1}

$$\nu_3 = 5.0576287736847894.$$

Konačno,

$$\lambda_3 = \frac{1}{\nu_3} + d_3 = 3.1977211149230786.$$

Normirani svojstveni vektori su (prema Mathematici)

$$u_1 = \begin{bmatrix} 0.8033295138152280 \\ 0.1840492832181065 \\ 0.07905323368406930 \\ 0.5608369993361552 \end{bmatrix}, \qquad u_2 = \begin{bmatrix} -0.4990210282687349 \\ 0.7111810660088434 \\ 0.1992737720673758 \\ 0.4533094621361650 \end{bmatrix},$$

$$u_3 = \begin{bmatrix} -0.1084443289271091 \\ -0.4327505232905481 \\ 0.8779704875654388 \\ 0.1735933036709976 \end{bmatrix}, \qquad u_4 = \begin{bmatrix} -0.3063976715031510 \\ -0.5225651512387552 \\ -0.4280284721537073 \\ 0.6706994532829344 \end{bmatrix}.$$

Matlab funkcija eig *daje*

$$u_1' = \begin{bmatrix} 0.8033295138152279 \\ 0.1840492832181065 \\ 0.07905323368406905 \\ 0.5608369993361553 \end{bmatrix}, \qquad u_2' = \begin{bmatrix} 0.4990210282687349 \\ -0.7111810660088435 \\ -0.1992737720673757 \\ -0.4533094621361650 \end{bmatrix},$$

$$u_3' = \begin{bmatrix} 0.1084443289271095 \\ 0.4327505232905485 \\ -0.8779704875654384 \\ -0.1735933036709980 \end{bmatrix}, \qquad u_4' = \begin{bmatrix} -0.3063976715031510 \\ -0.5225651512387552 \\ -0.4280284721537073 \\ 0.6706994532829344 \end{bmatrix}.$$

Provjera ortogonalnosti, uz $U_{eig} = [u_1'\ u_2'\ u_3'\ u_4']$*, daje sljedeći rezultat:*

$$U_{eig}^T \cdot U_{eig} = \begin{bmatrix} 1 & 5.551115 \cdot 10^{-17} & 1.665335 \cdot 10^{-16} & 5.551115 \cdot 10^{-17} \\ 5.551115 \cdot 10^{-17} & 1 & 2.775558 \cdot 10^{-17} & 2.498002 \cdot 10^{-16} \\ 1.665335 \cdot 10^{-16} & 2.775558 \cdot 10^{-17} & 1 & 5.551115 \cdot 10^{-17} \\ 5.551115 \cdot 10^{-17} & 2.498002 \cdot 10^{-16} & 5.551115 \cdot 10^{-17} & 1 \end{bmatrix}.$$

Algoritam aheig_basic *daje (bez naknadne ortogonalizacije)*

$$u_1'' = \begin{bmatrix} -0.8033295138152280 \\ -0.1840492832181065 \\ -0.07905323368406930 \\ -0.5608369993361552 \end{bmatrix}, \qquad u_2'' = \begin{bmatrix} 0.4990210282687349 \\ -0.7111810660088433 \\ -0.1992737720673757 \\ -0.4533094621361650 \end{bmatrix},$$

$$u_3'' = \begin{bmatrix} 0.1084443289271092 \\ 0.4327505232905481 \\ -0.8779704875654388 \\ -0.1735933036709976 \end{bmatrix}, \qquad u_4'' = \begin{bmatrix} 0.3063976715031510 \\ 0.5225651512387551 \\ 0.4280284721537074 \\ -0.6706994532829343 \end{bmatrix}.$$

Uz $U_{aheig\_basic} = [u_1'' \; u_2'' \; u_3'' \; u_4'']$ *i* $U_1 = U_{aheig\_basic}^T \cdot U_{aheig\_basic}$ *imamo*

$$U_1 = \begin{bmatrix} 1 & 5.551115 \cdot 10^{-17} & 0 & 0 \\ 5.551115 \cdot 10^{-17} & 1 & 2.775558 \cdot 10^{-17} & 0 \\ 0 & 2.775558 \cdot 10^{-17} & 1 & 8.326673 \cdot 10^{-17} \\ 0 & 0 & 8.326673 \cdot 10^{-17} & 1 \end{bmatrix}.$$

2.2 Točnost algoritma *aheig_basic*

U ovom odlomku analiziramo točnost algoritma *aheig_basic*. Pritom pretpostavljamo da su elementi matrice A definirane u (2.1) reprezentabilni u računalu i njih smatramo polaznim podacima. Za $\lambda \in \sigma(A)$ i njoj najbliži pol d_i, $i \in \{1, \ldots, n-1\}$, te

$$\mu = \lambda - d_i \in \sigma(A_i), \quad \nu = \frac{1}{\mu} \in \sigma(A_i^{-1})$$

uvodimo osnovne oznake:

matrica	egzaktna svojstvena vrijednost	izračunata svojstvena vrijednost
A	λ	$\widetilde{\lambda}$
A_i	μ	$-$
$\widetilde{A_i} = fl(A_i)$	$\hat{\mu}$	$\widetilde{\mu} = fl(\hat{\mu})$
A_i^{-1}	ν	$-$
$\widetilde{(A_i^{-1})} = fl(A_i^{-1})$	$\hat{\nu}$	$\widetilde{\nu} = fl(\hat{\nu})$

Pritom je

$$\widetilde{A_i} = fl(A_i) = \begin{bmatrix} D_1(I+E_1) & 0 & 0 & z_1 \\ 0 & 0 & 0 & \xi_i \\ 0 & 0 & D_2(I+E_2) & z_2 \\ z_1^T & \xi_i & z_2^T & a(1+\varepsilon_a) \end{bmatrix},$$

gdje su E_1 i E_2 dijagonalne matrice kojih su dijagonalni elementi po apsolutnoj vrijednosti omeđeni s ε_M, te $|\varepsilon_a| \leq \varepsilon_M$. Definiramo i $\kappa_\lambda, \kappa_\mu, \kappa_b \in \mathbb{R}$ s

$$\widetilde{\lambda} = fl(\lambda) = \lambda(1 + \kappa_\lambda \varepsilon_M), \tag{2.17}$$

$$\widetilde{\mu} = fl(\mu) = \mu(1 + \kappa_\mu \varepsilon_M), \tag{2.18}$$

$$\widetilde{b} = fl(b) = b(1 + \kappa_b \varepsilon_M) = b + \delta b. \tag{2.19}$$

Prvo pokazujemo vezu točnosti svojstvenih vrijednosti λ i μ, te kako točnost svojstvenog vektora pridruženog svojstvenoj vrijednosti λ (tj. μ) ovisi o točnosti μ. Nakon toga dajemo tvrdnje o točnosti svojstvene vrijednosti μ, koja ovisi o točnosti elemenata matrice A_i^{-1}, te o tome da li je ona svojstvena vrijednost matrice A_i koja je najbliža nuli. U ovom odlomku zbog preglednosti dajemo samo iskaze. Dokaze odgovarajućih tvrdnji dajemo u Dodatku.

Veza točnosti svojstvenih vrijednosti matrica A i A_i

Teorem 2.5 ([7, Teorem 4.1])**.** *Za κ_λ iz (2.17) i κ_μ iz (2.18) vrijedi*

$$|\kappa_\lambda| \leq \frac{|d_i| + |\mu|}{|\lambda|}(|\kappa_\mu| + 1).$$

Teorem 2.6 ([7, Korolar 4.1, 4.2])**.**

(i) Ako vrijedi i $\operatorname{sign}(d_i) = \operatorname{sign}(\mu)$ *ili* $d_i = 0$*, onda je*

$$\frac{|d_i| + |\mu|}{|\lambda|} = 1.$$

(ii) Ako se λ *nalazi između dva pola koji imaju isti predznak, onda je*

$$\frac{|d_i| + |\mu|}{|\lambda|} \leq 3.$$

Korolar 2.7 ([7, Korolar 4.5])**.** *Postoji najviše jedna svojstvena vrijednost* $\lambda' \in \sigma(A)$ *za koju ne vrijedi*

$$|\kappa_{\lambda'}| \leq 5(|\kappa_\mu| + 1).$$

Ako takva vrijednost postoji, onda je to svojstvena vrijednost matrice A najbliža nuli.

Sljedeću lemu dajemo bez dokaza.

Lema 2.8 ([6])**.** *Neka je zadan skup brojeva* $\{\hat{\lambda}_i\}_{i=1}^n$ *i neka je zadana dijagonalna matrica* $D = \operatorname{diag}(d_1, \ldots, d_{n-1})$*. Ako vrijedi*

$$\hat{\lambda}_1 > d_1 > \hat{\lambda}_2 > \cdots > d_{n-1} > \hat{\lambda}_n,$$

onda postoji streličasta matrica

$$\hat{H} = \begin{bmatrix} D & \hat{z} \\ \hat{z}^T & \hat{\alpha} \end{bmatrix}$$

čije su svojstvene vrijednosti $\{\hat{\lambda}_i\}_{i=1}^n$*. Vektor* $\hat{z}$ *i skalar* $\hat{\alpha}$ *su dani s*

$$|\hat{z}_i| = \sqrt{(d_i - \hat{\lambda}_1)(\hat{\lambda}_n - d_i) \prod_{j=1}^{i-1} \frac{(\hat{\lambda}_{j+1} - d_i)}{(d_j - d_i)} \prod_{j=i+1}^{n-1} \frac{(\hat{\lambda}_j - d_i)}{(d_j - d_i)}},$$

$$\hat{\alpha} = \hat{\lambda}_1 + \sum_{j=2}^{n} (\hat{\lambda}_j - d_{j-1}),$$

gdje za svako i, $i = 1, \ldots, n-1$*, predznak od* $\hat{z}_i$ *biramo po volji.*

Primjer 2.9. *Neka je*

$$D = [4, 3, -2, -4]^T, \quad \Lambda = [4.9, 3.1, -0.1, -2.2, -4.5]^T \in \mathbb{R}^5.$$

Pomoću leme 2.8 kreiramo streličastu matricu

$$\begin{bmatrix} D & \hat{z} \\ \hat{z}^T & \hat{\alpha} \end{bmatrix}, \quad \hat{z} = \begin{bmatrix} 1.9094976433606825 \\ 0.81013226433359986 \\ 0.74644155832858074 \\ 1.4072417931136476 \end{bmatrix}, \quad \hat{\alpha} = 0.2.$$

Tvrdnja korolara 2.7 vrijedi za sve svojstvene vrijednosti osim λ_3, za koju je $i = 3$, te vrijedi

$$\frac{|d_3| + |\mu_3|}{|\lambda_3|} = \frac{2.2 + 2.1}{0.1} = 43.$$

Dakle, svojstvena vrijednost λ_3 ne mora biti točno izračunata. Ako su svojstvene vrijednosti izračunate algoritmom aheig_basic dane u $\Lambda_{aheig\_basic} \in \mathbb{R}^5$, onda imamo

$$\Lambda - \Lambda_{aheig\_basic} = \begin{bmatrix} 0 \\ 0 \\ 3.053113 \\ 0 \\ 0 \end{bmatrix} \cdot 10^{-16},$$

tj. λ_3 nije točno izračunata. Također, ako s $\widetilde{\mu}_i$, $i = 1, \ldots, 5$ označimo izračunate svojstvene vrijednosti odgovarajućih pomaknutih matrica, imamo greške

$$\begin{bmatrix} \mu_1 - \widetilde{\mu}_1 \\ \mu_2 - \widetilde{\mu}_2 \\ \mu_3 - \widetilde{\mu}_3 \\ \mu_4 - \widetilde{\mu}_4 \\ \mu_5 - \widetilde{\mu}_5 \end{bmatrix} = \begin{bmatrix} 1.110223 \\ -0.693889 \\ 2.220446 \\ 1.387779 \\ 0 \end{bmatrix} \cdot 10^{-16}.$$

Dakle, greške u računanju svojstvenih vrijednosti pomaknutih matrica su manje od ε_M (u Matlabu, $\varepsilon_M = 2.2204 \cdot 10^{-16}$). To lijepo prikazuje ponašanje opisano u korolaru 2.7. Budući da svojstvene vektore računamo prema (2.14), i oni su po komponentama izračunati do na strojnu točnost i međusobno ortogonalni (to ćemo dokazati u sljedećem teoremu). Npr., za ilustraciju, matrica $U^T_{aheig\_basic} \cdot U_{aheig\_basic}$ je dana s

$$\begin{bmatrix} 1 & 6.938894 \cdot 10^{-18} & 5.555112 \cdot 10^{-17} & 2.775558 \cdot 10^{-17} & 2.775558 \cdot 10^{-17} \\ 6.938894 \cdot 10^{-18} & 1 & 2.775558 \cdot 10^{-17} & 2.428613 \cdot 10^{-17} & 6.938894 \cdot 10^{-18} \\ 5.555112 \cdot 10^{-17} & 2.775558 \cdot 10^{-17} & 1 & 5.555112 \cdot 10^{-17} & 5.555112 \cdot 10^{-17} \\ 2.775558 \cdot 10^{-17} & 2.428613 \cdot 10^{-17} & 5.555112 \cdot 10^{-17} & 1 & 2.775558 \cdot 10^{-17} \\ 2.775558 \cdot 10^{-17} & 6.938894 \cdot 10^{-18} & 5.555112 \cdot 10^{-17} & 2.775558 \cdot 10^{-17} & 1 \end{bmatrix}.$$

Napomena 2.10 (Invertiranje originalne matrice)**.** Za svojstvenu vrijednost matrice A koja je najmanja po modulu ne mora vrijediti ograda iz korolara 2.7 jer kvocijent

$$K_d(\lambda) \equiv \frac{|d_i| + |\mu|}{|\lambda|}$$

iz teorema 2.5 može biti 'velik'. To se dogodi ako je μ takav da je λ blizu nule, pa λ računamo kao razliku dvije bliske vrijednosti, te može doći do katastrofalnog kraćenja. Ta se svojstvena vrijednost može točno izračunati kao inverz po modulu najveće svojstvene vrijednosti metrice A^{-1} (npr. metodom bisekcije). Pritom je potrebno invertirati streličastu matricu A definiranu u (2.1). O tome daje informaciju sljedeća lema.

Lema 2.11. *Ako je matrica A definirana u (2.1) regularna, onda njen inverz ima sljedeći oblik*

$$A^{-1} = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \rho v v^T, \quad v = \begin{bmatrix} D^{-1}z \\ -1 \end{bmatrix}, \quad \rho = \frac{1}{\alpha - z^T D^{-1} z}. \tag{2.20}$$

Dokaz. Računamo

$$\begin{aligned} \begin{bmatrix} D & z \\ z^T & \alpha \end{bmatrix} \left(\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \rho v v^T \right) &= \begin{bmatrix} D & z \\ z^T & \alpha \end{bmatrix} \cdot \left(\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \rho \begin{bmatrix} D^{-1} z z^T D^{-1} & -D^{-1} z \\ -z^T D^{-1} & 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} I_{n-1} & 0 \\ z^T D^{-1} & 0 \end{bmatrix} + \rho \begin{bmatrix} z z^T D^{-1} - z z^T D^{-1} & -z + z \\ z^T D^{-1} z z^T D^{-1} - \alpha z^T D^{-1} & -z^T D^{-1} z + \alpha \end{bmatrix} \\ &= \begin{bmatrix} I_{n-1} & 0 \\ z^T D^{-1} & 0 \end{bmatrix} + \rho \begin{bmatrix} 0 & 0 \\ -\frac{1}{\rho} z^T D^{-1} & \frac{1}{\rho} \end{bmatrix} = I_n. \end{aligned}$$

□

U [1] je pokazano da su svojstvene vrijednosti matrice A^{-1} nultočke funkcije

$$f_{A^{-1}}(\lambda) = 1 + \rho \sum_{j=1}^{n-1} \frac{v_j^2}{d_j - \lambda} - \frac{\rho}{\lambda}.$$

Budući da je tražena svojstvena vrijednost po modulu najveća svojstvena vrijednost, ona će se točno izračunati (korolar 2.18), pa će i njen inverz biti točno izračunat. U računanju inverza A^{-1} vrijednost ρ ponekad treba računati s većom preciznošću. Ako se za nazivnik od ρ dobije 0, onda je matrica A numerički singularna, pa stavljamo $\lambda = 0$.

Točnost svojstvenih vektora

Svojstveni vektor računamo prema (2.14), pa njegova točnost ovisi o točnosti $\widetilde{\mu}$.

Teorem 2.12 ([7, Teorem 4.2]). *Neka vrijedi (2.18) i neka je*

$$\widetilde{x} = \begin{bmatrix} \widetilde{x_1} \\ \vdots \\ \widetilde{x_n} \end{bmatrix} = fl\left(\begin{bmatrix} D_1(I+E_1) - \widetilde{\mu}I)^{-1}z_1 \\ -\frac{\xi_i}{\widetilde{\mu}} \\ D_2(I+E_2) - \widetilde{\mu}I)^{-1}z_2 \\ -1 \end{bmatrix}\right) \tag{2.21}$$

izračunat nenormaliziran svojstveni vektor pridružen svojstvenoj vrijednosti λ odnosno svojstvenoj vrijednosti μ. Tada vrijedi

$$\widetilde{x}_j = x_j(1+\varepsilon_{x_j}), \quad |\varepsilon_{x_j}| \leq 3(|\kappa_\mu|+3)\varepsilon_M, \quad j = 1, \ldots, n.$$

Dakle, ako je κ_μ mali tj. ako je vrijednost $\widetilde{\mu}$ točno izračunata, onda će sve komponente nenormaliziranog svojstvenog vektora x biti izračunate s visokom relativnom točnošću. Budući da su greške računanja norme i skaliranja male, i normirani vektor $u = x/\|x\|_2$ će po komponentama biti izračunat s visokom relativnom točnošću. To povlači i da će normirani vektori biti međusobno ortogonalni.

Još trebamo pokazati kada će vrijednost $\widetilde{\mu}$ biti točno izračunata.

Točnost svojstvenih vrijednosti matrice $A_i = A - d_i$

Točnost A_i^{-1}

Teorem 2.13 ([7, Teorem 5.1]). *Neka je matrica A_i definirana relacijama (2.10) i (2.11). Tada vrijedi*

$$fl\left((A_i^{-1})_{jk}\right) = (A_i^{-1})_{jk}(1+\varepsilon_{jk}), \quad |\varepsilon_{jk}| \leq 3\varepsilon_M$$

za sve elemente matrice A_i^{-1} osim elementa b.

Dakle, svi elementi matrice A_i^{-1}, osim eventualno b, će biti izračunati s visokom relativnom točnosti. Sljedeći teorem govori o točnosti elementa b.

Teorem 2.14 ([7, Teorem 5.2, Korolar 5.1]). *Neka je matrica A_i definirana relacijama (2.10) i (2.11) i neka vrijedi (2.19). Tada je*

(i)

$$|\delta b| \leq \frac{1}{\zeta_i^2}(|a| + |z_1^T D_1 z_1| + |z_2^T D_2 z_2|)(n+5)\varepsilon_M,$$

(ii)

$$|\kappa_b| \leq (n+5)\frac{|a| + |z_1^T D_1 z_1| + |z_2^T D_2 z_2|}{|-a + z_1^T D_1 z_1 + z_2^T D_2 z_2|}. \tag{2.22}$$

Ocjena iz prethodnog teorema motivira definiciju veličine

$$K_b(\lambda) = \frac{|a| + |z_1^T D_1 z_1| + |z_2^T D_2 z_2|}{|-a + z_1^T D_1 z_1 + z_2^T D_2 z_2|}. \tag{2.23}$$

Primjer 2.15. *Neka je dana streličasta matrica*

$$\begin{bmatrix} 10^{10} & 0 & 0 & 10^{10} \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 10^{10} & 1 & 1 & 10^{10} \end{bmatrix}.$$

Svojstvene vrijednosti izračunate algoritmom aheig_basic su

$$\begin{aligned} \lambda_1 &= 2 \cdot 10^{10}, \\ \lambda_2 &= 2.26703509835452, \\ \lambda_3 &= 1.25865202250104, \\ \lambda_4 &= -0.5256871208375391. \end{aligned}$$

Računamo koeficijente K_b definirane u (2.23). Za svojstvenu vrijednost λ_1 imamo $i = 1$,

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 10^{10} \\ 0 & 2 - 10^{10} & 0 & 1 \\ 0 & 0 & 1 - 10^{10} & 1 \\ 10^{10} & 1 & 1 & 0 \end{bmatrix},$$

pa je

$$K_b(\lambda_1) = \frac{|\frac{1}{2-10^{10}} + \frac{1}{1-10^{10}}|}{|\frac{1}{2-10^{10}} + \frac{1}{1-10^{10}}|} = 1.$$

Dakle, po teoremu 2.14 znamo da je matrica A_1^{-1} točno izračunata. I svojstvena vrijednost λ_1 je točno izračunata (ostatak argumentacije za tu tvrdnju dajemo u nastavku odlomka). Za svojstvenu vrijdnost λ_2 je $i = 2$, pa imamo

$$A_2 = \begin{bmatrix} 10^{10}-2 & 0 & 0 & 10^{10} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \\ 10^{10} & 1 & 1 & 10^{10}-2 \end{bmatrix}$$

i

$$K_b(\lambda_2) = \frac{|10^{10}-2| + |\frac{10^{20}}{10^{10}-2}| + |-1|}{|2 - 10^{10} + \frac{10^{20}}{10^{10}-2} - 1|} \approx 10^{10}.$$

Slijedi da element b u matrici A_2^{-1} ne mora biti točno izračunat (i nije, izračuna se $b = 3.0$, a treba biti $b = 3.0000000004$, pa je i svojstvena vrijednost λ_2 netočna). Za svojstvene vrijednosti λ_3 i λ_4 je $i = 3$, pa je

$$A_3 = \begin{bmatrix} 10^{10}-1 & 0 & 0 & 10^{10} \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 10^{10} & 1 & 1 & 10^{10}-1 \end{bmatrix}$$

i

$$K_b(\lambda_3) = K_b(\lambda_4) = \frac{|10^{10}-1| + |\frac{10^{20}}{10^{10}-1} + 1|}{|1 - 10^{10} + \frac{10^{20}}{10^{10}-1} + 1|} \approx 10^{10}.$$

Dakle, ni element b u matrici A_3^{-1} ne mora biti točno izračunat (i nije, izračuna se $b = 3.0$, a treba biti $b = 3.0000000001$, pa su i svojstvene vrijednosti λ_3 i λ_4 netočne). Algoritam aheig_basic daje svojstvene vektore

$$u_1 = \begin{bmatrix} -0.7071067811865475 \\ -3.535533906286291 \cdot 10^{-11} \\ -3.535533906109514 \cdot 10^{-11} \\ -0.0.7071067811865475 \end{bmatrix}, \qquad u_2 = \begin{bmatrix} 0.2450963561190150 \\ -0.9178432257547516 \\ -0.1934408576224555 \\ -0.2450963560634508 \end{bmatrix},$$

$$u_3 = \begin{bmatrix} 0.2308354576071871 \\ 0.3113726139199682 \\ -0.8924556450247940 \\ -0.2308354575781330 \end{bmatrix}, \qquad u_4 = \begin{bmatrix} 0.6218060531511060 \\ 0.2461928273117208 \\ 0.4075580403683605 \\ -0.6218060531837936 \end{bmatrix},$$

od kojih je samo prvi točan. Posebno, ti svojstveni vektori nisu ni ortogonalni. Uz

$$U_{aheig\_basic} = [u_1\ u_2\ u_3\ u_4]$$

imamo

$$O_{aheig\_basic} \equiv \max |U^T_{aheig\_basic} U_{aheig\_basic} - I_4| = 1.683195 \cdot 10^{-11}.$$

Ako se vrijednost b izračuna korištenjem četverostruke preciznosti, te tada zaokruži na dvostruku preciznost (to zovemo algoritam aheig_quad i dajemo kao Algoritam 6), onda je ta vrijednost točno izračunata, te su i svojstvene vrijednosti točno izračunate i iznose

$$\begin{aligned}
\lambda'_1 &= 2 \cdot 10^{10}, \\
\lambda'_2 &= 2.267035098330492, \\
\lambda'_3 &= 1.258652022495711, \\
\lambda'_4 &= -0.5256871208762033.
\end{aligned}$$

Svojstveni vektori

$$u'_1 = \begin{bmatrix} -0.7071067811865475 \\ -3.535533906286291 \cdot 10^{-11} \\ -3.535533906109514 \cdot 10^{-11} \\ -0.0.7071067811865475 \end{bmatrix}, \qquad u'_2 = \begin{bmatrix} 0.2450963561002613 \\ -0.9178432257671137 \\ -0.1934408576113228 \\ -0.2450963560446971 \end{bmatrix},$$

$$u'_3 = \begin{bmatrix} 0.2308354576035603 \\ 0.3113726139128379 \\ -0.8924556450291578 \\ -0.2308354575745062 \end{bmatrix}, \qquad u'_4 = \begin{bmatrix} 0.6218060531543005 \\ 0.2461928273092168 \\ 0.4075580403601259 \\ -0.621806053186988 \end{bmatrix},$$

su također točno izračunati i ortogonalni. Uz

$$U_{aheig\_quad} = [u'_1\ u'_2\ u'_3\ u'_4]$$

vrijedi

$$O_{aheig\_quad} \equiv \max |U^T_{aheig\_quad} U_{aheig\_quad} - I_4| = 1.457168 \cdot 10^{-16}.$$

Napomena 2.16. Sada ćemo ilustrirati zašto je četverostruka preciznost uvijek dovoljna kako bi osigurali da nema 'bitnog' skraćivanja u b. Prvo uvodimo veličine P i Q.

(i) Ako je $a < 0$, definiramo

$$P = -a + \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i} > 0,$$
$$-Q = \sum_{k=i+1}^{n} \frac{\zeta_k^2}{d_k - d_i} < 0.$$

(ii) Ako je $a > 0$, definiramo

$$P = \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i} > 0,$$
$$-Q = -a + \sum_{k=i+1}^{n} \frac{\zeta_k^2}{d_k - d_i} < 0.$$

Propozicija 2.17. *Tada je*

$$fl(P) = P(1 + \varepsilon_P), \qquad |\varepsilon_P| \leq (n+2)\varepsilon_M,$$
$$fl(Q) = Q(1 + \varepsilon_Q), \qquad |\varepsilon_Q| \leq (n+2)\varepsilon_M.$$

Dokaz. Ako zanemarujemo članove reda $O(\varepsilon_M^2)$ i više, za $k \in \{1, \ldots, n\}, k \neq i$ imamo

$$fl(\frac{\zeta_k^2}{d_k - d_i}) = \frac{\zeta_k^2(1+\varepsilon_1)}{(d_k - d_i)(1+\varepsilon_2)}(1+\varepsilon_3) = \frac{\zeta_k^2}{d_k - d_i}(1 + \varepsilon_1 - \varepsilon_2 + \varepsilon_3)$$
$$= \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_k), \quad |\varepsilon_k| \leq 3\varepsilon_M.$$

U sumi

$$\sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}$$

imamo $(i-1)$ član, pa se zbrajanje vrši $(i-2)$ puta. Slijedi da je

$$fl\left(\sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}\right) = \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_4), \quad |\varepsilon_4| \leq (i+1)\varepsilon_M.$$

Analogno,

$$fl\left(\sum_{k=i+1}^{n-1} \frac{\zeta_k^2}{d_k - d_i}\right) = \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_5), \quad |\varepsilon_5| \leq (n-i+1)\varepsilon_M.$$

U svakom slučaju vrijedi

$$\begin{aligned} fl(P) &= P(1+\varepsilon_P), \qquad |\varepsilon_P| \le (n+2)\varepsilon_M, \\ fl(Q) &= Q(1+\varepsilon_Q), \qquad |\varepsilon_Q| \le (n+2)\varepsilon_M. \end{aligned}$$

□

Dakle, promatramo stabilnost računa

$$(P(1+\varepsilon_P) - Q(1+\varepsilon_Q))(1+\varepsilon_1) = (P-Q)(1+\varepsilon_{PQ}), \quad P, Q > 0.$$

Pretpostavimo prvo da je $P \neq Q$, $fl(P) \neq fl(Q)$. Tada imamo

$$\varepsilon_{PQ} = \frac{P(\varepsilon_P + \varepsilon_1) - Q(\varepsilon_Q + \varepsilon_1)}{P - Q}$$

i

$$|\varepsilon_{PQ}| \le \frac{P|\varepsilon_P + \varepsilon_1| + Q|\varepsilon_Q + \varepsilon_1|}{|P-Q|} \le \frac{(P+Q)(n+3)\varepsilon_M}{|P-Q|}.$$

Označimo

$$P = Q + x.$$

Tada je

$$\frac{P+Q}{|P-Q|} = \frac{2Q+x}{|x|} = \frac{2Q}{|x|} + \text{sign}(x).$$

Vidi se da je najgori mogući slučaj je kad se $fl(P)$ i $fl(Q)$ razlikuju tek na zadnjoj znamenci. Ako je $|x| \approx Q \cdot \varepsilon_M$, onda imamo

$$\frac{P+Q}{|P-Q|} = \frac{Q(2+\varepsilon_M)}{Q \cdot \varepsilon_M} \le \frac{2}{\varepsilon_M}.$$

Ako računamo u standardnoj preciznosti ($\varepsilon_M \approx 10^{-16}$), u tom slučaju imamo

$$|\varepsilon_{PQ}| \le \frac{(P+Q)(n+3)\varepsilon_M}{|P-Q|} \le \frac{2}{\varepsilon_M}(n+3)\varepsilon_M = 2(n+3).$$

Računanje u četverostrukoj preciznosti (s $\varepsilon_M^2 \approx 10^{-32}$) daje

$$|\varepsilon_{PQ}| \le \frac{2}{\varepsilon_M}(n+3)\varepsilon_M^2 = 2(n+3)\varepsilon_M.$$

Ako je $fl(P) = fl(Q)$ tj. $P(1+\varepsilon_P) = Q(1+\varepsilon_Q)$, onda je $fl(b) = 0$, a za b vrijedi

$$b = P - Q = Q\varepsilon_Q - P\varepsilon_P,$$

pa je

$$|b| \leq (P+Q)(n+2)\varepsilon_M.$$

Ako je $P = Q$, onda je $b = 0$, a za $fl(b)$ vrijedi

$$fl(b) = P\varepsilon_P - Q\varepsilon_Q,$$

pa je

$$|fl(b)| \leq (P+Q)(n+2)\varepsilon_M.$$

Dakle, četverostruka preciznost je dovoljna za točno izračunati b.

Matlab funkcija *aheig_quad* (Algoritam 6):

ULAZ: dijagonala D, vektor z, skalar α, indeks tražene svojstvene vrijednosti k.

24 Umjesto Matlab funkcije *invA* pozivamo Matlab funkciju *invA_quad* koja veličinu b iz (2.12) računa korištenjem četverostruke preciznosti.

IZLAZ: svojstvena vrijednost λ_k i pripadni normirani svojstveni vektor u_k.

Matlab funkcija *invA_quad* (Algoritam 6):

ULAZ: dijagonala D, vektor z, skalar α, indeks najbližeg pola i.

11-15 Za danu matricu

$$A = \begin{bmatrix} diag(D) & z \\ z^T & \alpha \end{bmatrix}$$

i indeks $i \in \{1, \ldots, n-1\}$ računamo element b matrice A_i^{-1}, dan u (2.13), korištenjem četverostruke preciznosti. Koristimo Matlab naredbe *vpa* i *digits*.

11 Povećamo preciznost na 32 značajne znamenke korištenjem naredbe *digits*.

12-14 Računamo veličinu b korištenjem aritmetike varijabilne preciznosti. Nakon što smo postavili broj značajnih znamenki na 32, naredbom *vpa* sve elemente računamo sa 32 točne znamenke.

15 Pomoću naredbe *digits* vratimo preciznost na 16 značajnih znamenki.

IZLAZ: dijagonala matrica D_1^{-1}, D_2^{-1}, vektori w_1, $w2$, skalari $w_\xi = 1/\zeta_i$, b

Algoritam 6

```
function [lambda,u] = aheig_quad(D,z,alpha,k)
n = length(D)+1;
if k == 1
    shift = D(1);
    i = 1;
    side = 'R';
elseif k == n
    shift = D(n-1);
    i = n-1;
    side = 'L';
else
    middle = (D(k-1)+D(k))/2;
    Fmiddle = alpha - middle - sum(z.^2./(D-middle));
    if Fmiddle < 0
        shift = D(k);
        i = k;
        side = 'R';
    else
        shift = D(k-1);
        i = k-1;
        side = 'L';
    end
end
[invD1,invD2,w1,w2,wKsi,b] = invA_quad(D,z,alpha,i);
lambda = bisect([invD1;0;invD2],[w1;wKsi;w2],b,side);
lambda = 1/lambda;
u = vect(D-shift,z,lambda);
lambda = lambda + shift;
end
```

```
function [invD1,invD2,w1,w2,wKsi,b] = invA_quad(D,z,alpha,i)
n = length(D)+1;
a = alpha-D(i);
D = D-D(i);
invD1 = 1./D(1:i-1);
invD2 = 1./D(i+1:n-1);
wKsi = 1/z(i);
w1 = -z(1:i-1).*invD1*wKsi;
w2 = -z(i+1:n-1).*invD2*wKsi;

digits(32)
a = vpa(a); z = vpa(z); D = vpa(D);
b = -a+sum(z(1:i-1).^2./D(1:i-1))+sum(z(i+1:n-1).^2./D(i+1:n-1))/z(i)^2;
b = double(vpa(b));
digits(16)
end
```

Točnost bisekcije

Neka je matrica A definirana kao u relaciji (2.1) i neka je $\lambda_{\max} \in \sigma(A)$ po modulu najveća svojstvena vrijednost od A tj.

$$|\lambda_{\max}| = \max\{|\lambda|,\ \lambda \in \sigma(A)\}, \tag{2.24}$$

a $\widetilde{\lambda}_{\max}$ najveća svojstvena vrijednost izračunata metodom bisekcije.

Korolar 2.18 ([7, Korolar 4.8])**.** *Neka je $\widetilde{\lambda}_{max}$ izračunata metodom bisekcije za matricu A. Tada je*

$$\widetilde{\lambda}_{max} = \lambda_{max}(1 + k_{\lambda_{max}}\varepsilon_M), \quad |k_{\lambda_{max}}| \leq 1.06n(\sqrt{n} + 1).$$

Slična ocjena vrijedi i za svojstvene vrijednosti koje su reda veličine $|\lambda_{\max}|$.

Korolar 2.19 ([7, Korolar 4.9])**.** *Neka je $\lambda_k \in \sigma(A)$ i neka je $\widetilde{\lambda}_k$ izračunata metodom bisekcije za matricu A. Ako je $\frac{|\lambda_{max}|}{|\lambda_k|} = s$, onda vrijedi*

$$\widetilde{\lambda}_k = \lambda_k(1 + k_{\lambda_k}\varepsilon_M), \quad |k_{\lambda_k}| \leq s \cdot 1.06n(\sqrt{n} + 1).$$

Neka je A_i^{-1} matrica iz (2.12) i neka je ν njena svojstvena vrijednost za koju vrijedi

$$|\nu_{\max}| = \max\{|\nu|,\ \nu \in \sigma(A_i^{-1})\}. \tag{2.25}$$

Omjer $\nu_{\max}$ i svojstvene vrijednosti koju metodom bisekcije računamo iz matrice A_i^{-1} označavamo

$$K_\nu(\lambda) = \frac{|\nu_{\max}|}{\nu_i}. \tag{2.26}$$

Točnost svojstvenih vrijednosti matrice A_i

Teorem 2.20 ([7, Korolar 5.4])**.** *Neka je A_i^{-1} matrica iz (2.12), ν_{max} njena po modulu najveća svojstvena vrijednost definirana u 2.25 te neka je*

$$\hat{\nu}_{max} = \nu_{max}(1 + k_{\nu_{max}}\varepsilon_M), \quad |k_{\nu_{max}}\varepsilon_M| < 1$$

odgovarajuća svojstvena vrijednost matrice $\widetilde{(A_i^{-1})}$. Ako uz oznake iz teorema 2.14 vrijedi

$$|\kappa_b| \leq C,$$

onda je

$$|k_{\nu_{max}}| \leq \min\{3\sqrt{n} + \frac{(n+5)}{|\nu_{max}|} \cdot \frac{|a| + |z_1^T D_1^{-1} z_1| + |z_2^T D_2^{-1} z_2|}{\zeta_i^2},\ \sqrt{n}\max\{3, C\}\}.$$

Ocjena (2.22) iz teorema 2.14 povlači

$$|k_{\nu_{max}}| \leq \cdot \min\{3\sqrt{n} + \frac{(n+5)}{|\nu_{max}|} \cdot \frac{|a| + |z_1^T D_1^{-1} z_1| + |z_2^T D_2^{-1} z_2|}{\zeta_i^2},\ \sqrt{n}\max\{1, (n+5)K_b(\lambda)\}\}.$$

Prethodni korolar motivira definiciju veličine

$$K_\mu(\lambda) = |\mu| \cdot \frac{|a| + |z_1^T D_1^{-1} z_1| + |z_2^T D_2^{-1} z_2|}{\zeta_i^2}. \tag{2.27}$$

Teorem 2.21 ([7, Korolar 5.5])**.** *Uz uvjete iz teorema 2.20 vrijedi*

$$|k_{\nu_{max}}| \leq 6 \cdot (\sqrt{n} + (n-2)\frac{1}{\zeta_i} \max_{\substack{k=1,\ldots,n-1 \\ k\neq i}} |\zeta_k|).$$

Lema 2.22 ([7, Lema 5.2])**.** *Neka vrijede uvjeti iz teorema 2.20 i neka je* $\mu = \frac{1}{\nu_{max}}$*, a* $\hat{\mu}$ *svojstvena vrijednost matrice* $\widetilde{A}_i$*. Tada vrijedi*

$$\hat{\mu} = \mu(1 + k_\mu \varepsilon_M), \quad |k_\mu| \leq |k_{\nu_{max}}| + 1.$$

Napomena 2.23. Još treba promotriti što se događa kada svojstvena vrijednost ν koju računamo nije reda veličine po modulu najveće svojstvene vrijednosti matrice A_i^{-1} tj. kada za veličinu K_ν, definiranu u (2.26), vrijedi $K_\nu \gg 1$. Razlikujemo dva slučaja:

(i) Vrijednost K_ν je velika za rubne svojstvene vrijednosti, λ_1 ili λ_n. Tada te svojstvene vrijednosti možemo izračunati bisekcijom na originalnoj matrici A. Točnost po modulu veće od tih svojstvenih vrijednosti garantira korolar 2.18, točnost druge rubne svojstvene vrijednosti ovisi o faktoru s iz korolara 2.19. Algoritam u kojem se rubne svojstvene vrijednosti računaju bisekcijom iz originalne matrice A zovemo *aheig_A* i dajemo u Algoritmu 7.

(ii) Vrijednost K_ν je velika za neku od svojstvenih vrijednosti $\lambda_1, \ldots, \lambda_{n-1}$. Za te 'unutarnje' svojstvene vrijednosti točnost nije garantirana, ali eksperimenti pokazuju da svojstvene vrijednosti ne odlutaju daleko od one koja je tom polu najbliža jer su ograničene susjednim polom.

Algoritam 7

```
1  function [lambda,u] = aheig_A(D,z,alpha,k)
2  n = length(D)+1;
3  if k == 1
4      side = 'R';
5  elseif k == n
6      side = 'L';
7  end
8  lambda = bisect(D,z,alpha,side);
9  u = vect(D,z,lambda);
10 end
```

Matlab funkcija *aheig_A* (Algoritam 7):

ULAZ: dijagonala D, vektor z, skalar α, indeks tražene svojstvene vrijednosti k.

3-7 Znamo da mora vrijediti $k \in \{1, n\}$. Ako je $k = 1$ tj. tražimo najveću svojstvenu vrijdnost matrice A, onda je $\lambda_k = \lambda_1 > d_1$, pa se λ_k na brojevnom pravcu nalazi desno od d_1, te stavljamo $side = 'R'$. Ako je $k = n$ tj. tražimo najmanju svojstvenu vrijednost matrice A, onda je $\lambda_k = \lambda_n < d_{n-1}$, pa stavimo $side = 'L'$.

8 Tražimo nultočku funkcije f_A metodom bisekcije. Ako je $side = 'R'$, dobivamo najveću odnosno najdesniju nultočku, a ako je $side = 'L'$, onda dobivamo najmanju odnosno najljeviju nultočku funkcije f_A.

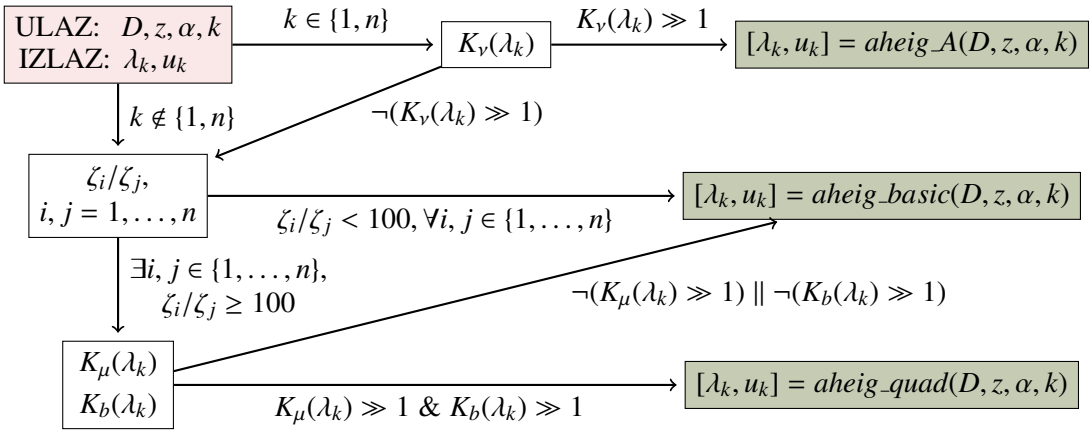
9 Vektor u_k računamo kao $u_k = x_k/\|x_k\|_2$, gdje je x_k dan s (2.9).

IZLAZ: svojstvena vrijednost λ_k i pripadni normirani svojstveni vektor u_k.

2.3 Algoritam *aheig*

Rezultate prethodnog odjeljka možemo sažeti na sljedeći način (Slika 2.2):

- Ako je $K_\nu(\lambda_k) \gg 1$, $k \in \{1, n\}$, svojstvenu vrijednost možemo izračunati metodom bisekcije na originalnoj matrici A, a svojstveni vektor pomoću (2.9) tj. koristimo algoritam *aheig_A* (Napomena 2.23, teorem 2.12). a svojstveni vektor pomoću (2.9) tj. koristimo algoritam *aheig_invA* (Napomena 2.10, korolar 2.18, teorem 2.12).
- Ako vrijedi $\zeta_i/\zeta_j < 100, \forall i, j \in \{1, \ldots, n\}$, koristimo algoritam *aheig_basic* (teorem 2.21, lema 2.22, teorem 2.12, korolar 2.7).
- Ako je je jedna od veličina $K_b(\lambda_k)$ i $K_\mu(\lambda_k)$ reda veličine $O(1)$, koristimo algoritam *aheig_basic* (teorem 2.20, lema 2.22, teorem 2.12, korolar 2.7).
- Ako je $\max\{K_b(\lambda_k), K_\mu(\lambda_k)\} \gg 1$, veličinu b računamo s četverostrukom preciznošću tj. koristimo algoritam *aheig_quad* (Napomena 2.16, teorem 2.20, teorem 2.12, korolar 2.7).



Slika 2.2: Algoritam *aheig*.

2.4 Algoritam *aheig* za matrice reda 3 i 4

U ovom odjeljku opisujemo primjenu algoritma *aheig* na matrice reda 3 i 4, što zovemo *aheig*3 i *aheig*4. Prilažemo Matlab kodove tih algoritama, te opisujemo svrhu pojedinih linija koda.

Matlab funkcija *aheig3* (Algoritam 8):

ULAZ: dijagonala D, vektor z, skalar α.

3-4 Određujemo pol najbliži svojstvenoj vrijednosti λ_2 pomoću funkcije f_A definirane u 2.6.

5-9 Ako vrijedi $f_A(d_1 + d_2)/2 > 0$, onda je najbliži pol d_1.

6 Svojstvene vrijednosti μ_1 i μ_2 računamo metodom bisekcije za istu matricu A_1^{-1}, pa nju računamo samo jednom.

7-8 Za računanje svojstvene vrijednosti μ_3 je potrebno izračunati matricu A_2^{-1}, te na kraju ispermutirati retke izračunatog vektora.

9 Računamo $\lambda_i = \mu_i + d_1$, za $i = 1, 2$.

11-14 Ako vrijedi $f_A(d_1 + d_2)/2 < 0$, onda je najbliži pol d_2.

11 Za računanje svojstvene vrijednosti μ_1 je potrebno izračunati matricu A_1^{-1}.

12-13 Svojstvene vrijednosti μ_2 i μ_3 računamo metodom bisekcije za istu matricu A_2^{-1}, pa nju računamo samo jednom. Na kraju ispermutiramo retke izračunatih svojstvenih vektora.

14 Računamo $\lambda_i = \mu_i + d_1$, za $i = 1, 2$.

IZLAZ: vektor $E = [\lambda_1, \lambda_2, \lambda_3]$, matrica $U = [u_1, u_2, u_3]$.

Algoritam 8

```
1  function [E,U]=aheig3(D,z,alpha)
2  E = zeros(3,1); U = zeros(3);
3  middle = (D(1)+D(2))/2;
4  Fmiddle = alpha - middle - sum(z.^2./(D-middle));
5  if Fmiddle > 0
6      [E(1:2),U(:,1:2)]=fja3(D(1),D(2),z(1),z(2),alpha,'R',2);
7      [E(3),U1] = fja3(D(2),D(1),z(2),z(1),alpha,'L',1);
8      U(:,3) = [U1(2,:); U1(1,:); U1(3,:)];
9      E = E + [D(1); D(1); D(2)];
10 else
11     [E(1),U(:,1)]=fja3(D(1),D(2),z(1),z(2),alpha,'R',1);
12     [E(2:3),U1]=fja3(D(2),D(1),z(2),z(1),alpha,'R',2);
13     U(:,2:3) = [U1(2,:); U1(1,:); U1(3,:)];
14     E = E + [D(1); D(2); D(2)];
15 end
16 end
```

Algoritam 9

```
function [E,U] = fja3(d,dd,z,zz,alpha,side,num)
E = zeros(num,1); U = zeros(3,num);
a = alpha - d; d1 = dd-d; invDD = 1/d1;
wKsi = 1/z; ww = -zz*invDD*wKsi;
b = (-a + zz*zz*invDD)*wKsi*wKsi;
E(1) = bisekcija3(invDD, ww, wKsi, b, side);
if (z/zz ≤ 0.01 || z/zz ≥ 100)
 tmp = abs(a) + abs(zz*zz*dd);
 Kb = tmp/(-a + zz*zz*dd);
 if Kb≥100
    Kmi = E(1)*tmp/(z*z);
    if Kmi≥100
         digits(32);
         a = vpa(a); z = vpa(z); zz = vpa(zz); d1 = vpa(d1);
         b = (-a + zz*zz/d1)/(z*z);
         b = double(vpa(b));
         digits(16);
         E(1) = bisekcija3(invDD, ww, wKsi, b, side);
    end
 end
end
U(:,1) = [-z/E(1); zz/(d1-E(1)); -1];
U(:,1) = U(:,1)/norm(U(:,1),2);
if num>1
    E(2) = bisekcija3(invDD, ww, wKsi, b, 'L');
    U(:,2) = [-z/E(2); zz/(d1-E(2)); -1];
    U(:,2) = U(:,2)/norm(U(:,2),2);
end
end
```

```
function lambda = bisekcija3(invDD, ww, wKsi, b, side)
if side == 'L'
    left = min(min(invDD-abs(ww),-abs(wKsi)),b-abs(ww)-abs(wKsi));
    right = min(invDD,0);
else
    left = max(invDD,0);
    right = max(max(invDD+abs(ww),abs(wKsi)),b+abs(ww)+abs(wKsi));
end
middle = (left+right)/2;
while(right-left)>2*eps*max(abs(left),abs(right))
    Fmiddle = b - middle - ww*ww/(invDD-middle)+wKsi*wKsi/middle;
    if Fmiddle>0
        left = middle;
    else
        right = middle;
    end
    middle =(left+right)/2;
end
lambda = 1/middle;
end
```

Matlab funkcija *fja3* (Algoritam 9):

ULAZ: Elementi matrice A: $d = d_i$, dd je preostali dijagonalni element od A, $z = \zeta_i$, zz je preostali element vektora $[\zeta_1, \zeta_2]$, $alpha$, varijabla *side* za metodu bisekcije, broj svojstvenih vrijednosti koje tražimo pomoću matrice A_i^{-1}: $num \in \{1, 2\}$.

3-4 Računamo elemente matrice A_i^{-1} prema (2.11), (2.12) i (2.13).

9-19 Izračunamo veličinu μ metodom bisekcije za matricu A_i^{-1}, pri čemu je element b izračunat s dvostrukom preciznošću.

9 Ako vrijedi $0.01 < z/zz < 100$, onda je prema teoremu 2.21 veličina b iz (2.13) točno izračunata.

11-22 Računamo vrijednosti veličina K_b i K_μ prema (2.23) i (2.27), respektivno. Ako su obje vrijednosti velike potrebno je ponovo izračunati b, ali s četverostrukom preciznošću, te ponoviti postupak računanja svojstvene vrijednosti μ metodom bisekcije iz točno izračunate matrice A_i^{-1}.

23-24,28-29 Računamo odgovarajuće svojstvene vektore prema (2.14), te ih normiramo.

IZLAZ: tražene svojstvene vrijednosti u vektoru $E \in \mathbb{R}^{num}$ i pripadni normirani svojstveni vektori u matrici $U \in \mathbb{R}^{3\times num}$.

Matlab funkcija *aheig4* (Algoritam 10):

ULAZ: dijagonala D, vektor z, skalar α.

2-3 Pomoću funkcije f_A definirane u 2.6 određujemo najbliže polove svojstvenim vrijednostima λ_2 i λ_3.

8 Vrijedi $f_A(d_1 + d_2)/2 > 0$, pa je d_1 najbliži pol svojstvenoj vrijednosti λ_2. Zato svojstvene vrijednosti μ_1 i μ_2 računamo metodom bisekcije za istu matricu A_1^{-1}, a nju računamo samo jednom.

10-14 Vrijedi $f_A(d_2 + d_3)/2 > 0$, pa je d_2 najbliži pol svojstvenoj vrijednosti λ_3. Svojstvenu vrijednosti μ_3 računamo pomoću matrice A_2^{-1}, a μ_4 pomoću A_3^{-1}.

16-17 Vrijedi $f_A(d_2 + d_3)/2 < 0$, pa je d_3 najbliži pol svojstvenoj vrijednosti λ_3. Zato svojstvene vrijednosti μ_3 i μ_4 računamo bisekcijom na matrici A_3^{-1}.

21 Vrijedi $f_A(d_1 + d_2)/2 < 0$, pa je d_2 najbliži pol svojstvenoj vrijednosti λ_2. Svojstvenu vrijednost μ_1 računamo bisekcijom na matrici A_1^{-1}.

23-26 Vrijedi $f_A(d_2 + d_3)/2 > 0$, pa je d_2 najbliži pol i svojstvenoj vrijednosti λ_3, te vrijednosti μ_2 i μ_3 računamo bisekcijom na matrici A_2^{-1}. Vrijednost μ_4 računamo pomoću matrice A_3^{-1}.

29-32 Vrijedi $f_A(d_2 + d_3)/2 < 0$, pa je d_3 najbliži pol svojstvenoj vrijednosti λ_2. Vrijednosti μ_3 i μ_4 računamo bisekcijom na matrici A_3^{-1}. Vrijednost μ_2 računamo pomoću matrice A_2^{-1}.

13 Računamo $\lambda_k = \mu_k + d_i$.

IZLAZ: vektor $E = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]$, matrica $U = [u_1, u_2, u_3, u_4]$.

Algoritam 10

```
function [E,U]=aheig4(D,z,alpha)
E = zeros(4,1); U = zeros(4);
middle = (D(1)+D(2))/2;
Fmiddle2 = alpha - middle - sum(z.^2./(D-middle));
middle = (D(2)+D(3))/2;
Fmiddle3 = alpha - middle - sum(z.^2./(D-middle));
if Fmiddle2 > 0
    [E(1:2),U(:,1:2)]=fja4(D(1),D(2:3),z(1),z(2:3),alpha,'R',2);
    if Fmiddle3 > 0
      [E(3),U1] = fja4(D(2),D(1:2:3),z(2),z(1:2:3),alpha,'L',1);
      U(:,3) = [U1(2,:); U1(1,:); U1(3:4,:)];
      [E(4),U1] = fja4(D(3),D(1:2),z(3),z(1:2),alpha,'L',1);
      U(:,4) = [U1(2:3,:); U1(1,:); U1(4,:)];
      E = E + [D(1); D(1); D(2); D(3)];
    else
      [E(3:4),U1]=fja4(D(3),D(1:2),z(3),z(1:2),alpha,'R',2);
      U(:,3:4) = [U1(2:3,:); U1(1,:); U1(4,:)];
      E = E + [D(1); D(1); D(3); D(3)];
    end
else
    [E(1),U(:,1)]=fja4(D(1),D(2:3),z(1),z(2:3),alpha,'R',1);
    if Fmiddle3 > 0
      [E(2:3),U1]=fja4(D(2),D(1:2:3),z(2),z(1:2:3),alpha,'R',2);
      U(:,2:3) = [U1(2,:); U1(1,:); U1(3:4,:)];
      [E(4),U1] = fja4(D(3),D(1:2),z(3),z(1:2),alpha,'L',1);
      U(:,4) = [U1(2:3,:); U1(1,:); U1(4,:)];
      E = E + [D(1); D(2); D(2); D(3)];
    else
      [E(2),U1]=fja4(D(2),D(1:2:3),z(2),z(1:2:3),alpha,'R',1);
      U(:,2) = [U1(2,:); U1(1,:); U1(3:4,:)];
      [E(3:4),U1] = fja4(D(3),D(1:2),z(3),z(1:2),alpha,'R',2);
      U(:,3:4) = [U1(2:3,:); U1(1,:); U1(4,:)];
      E = E + [D(1); D(2); D(3); D(3)];
end
end
```

Algoritam 11

```
function [E,U] = fja4(d,dd,z,zz,alpha,side,num)
E = zeros(num,1); U = zeros(4,num);
a = alpha - d; d1 = dd-d*ones(2,1); invDD = 1./d1;
wKsi = 1/z; ww = -zz.*invDD*wKsi;
b = (-a + sum(zz.*zz.*invDD))*wKsi*wKsi;
E(1) = bisekcija4(invDD, ww, wKsi, b, side);
if (min(z*ones(2,1)./zz) ≤ 0.01 || max(z*ones(2,1)./zz) ≥ 100)
 tmp = abs(a) + abs(sum(zz.*zz.*dd));
 Kb = tmp/(-a + sum(zz.*zz.*dd));
 if Kb≥100
    Kmi = E(1)*tmp/(z*z);
    if Kmi≥100
         digits(32);
         a = vpa(a); z = vpa(z); zz = vpa(zz); d1 = vpa(d1);
         b = (-a + zz.*zz./d1)/(z*z);
         b = double(vpa(b));
         digits(16);
         E(1) = bisekcija4(invDD, ww, wKsi, b, side);
    end
 end
end
U(:,1) = [-z/E(1); zz./(d1-E(1)*ones(2,1)); -1];
U(:,1) = U(:,1)/norm(U(:,1),2);
if num>1
    E(2) = bisekcija4(invDD, ww, wKsi, b, 'L');
    U(:,2) = [-z/E(2); zz./(d1-E(2)*ones(2,1)); -1];
    U(:,2) = U(:,2)/norm(U(:,2),2);
end
end
```

```
function lambda = bisekcija4(invDD, ww, wKsi, b, side)
if side == 'L'
    left = min([invDD-abs(ww); -abs(wKsi); b-norm(ww,1)-abs(wKsi)]);
    right = min([invDD; 0]);
else
    left = max([invDD; 0]);
    right = max([invDD+abs(ww); abs(wKsi); b+norm(ww,1)+abs(wKsi)]);
end
middle = (left+right)/2;
while(right-left)>2*eps*max(abs(left),abs(right))
    Fmiddle = b - middle - ...
        sum(ww.^2./(invDD-middle*ones(2,1)))+wKsi*wKsi/middle;
    if Fmiddle>0
        left = middle;
    else
        right = middle;
    end
    middle =(left+right)/2;
end
lambda = 1/middle;
end
```

Poglavlje 3

Jacobijeva rotacija

Neka je $A = [a_{ij}]_{i,j=1}^{n} \in \mathbb{R}^{n\times n}$ simetrična matrica i neka je $1 \leq p < q \leq n$, $a_{pq} \neq 0$. Izvandijagonalne elemente a_{pq} i a_{qp} možemo poništiti korištenjem ravninske rotacije dane matricom

$$J_{(p,q)} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos\varphi & & \sin\varphi & & \\ & & & 1 & & & \\ & & \vdots & & \vdots & & \\ & & & & & 1 & \\ & & -\sin\varphi & & \cos\varphi & & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{3.1}$$

Matrica $J_{(p,q)}$ na dijagonali ima jedinice, osim na mjestima $j_{pp} = j_{qq} = \cos\varphi$, a izvan dijagonale nule, osim na $j_{pq} = \sin\varphi$, $j_{qp} = -\sin\varphi$. Kut φ mora biti takav da za matricu

$$J_{(p,q)}^{T} A J_{(p,q)} = [a'_{rs}]_{r,s=1}^{n} \equiv A' \tag{3.2}$$

vrijedi

$$a'_{pq} = a'_{qp} = 0. \tag{3.3}$$

Vrijedi

$$a'_{pq} = a'_{qp} = a_{pq}(\cos^2\varphi - \sin^2\varphi) + (a_{pp} - a_{qq})\sin\varphi\cos\varphi, \tag{3.4}$$
$$a'_{rp} = a_{rp}\cos\varphi - a_{rq}\sin\varphi,\ r \neq p, q \tag{3.5}$$
$$a'_{rq} = a_{rq}\cos\varphi + a_{rp}\sin\varphi,\ r \neq p, q \tag{3.6}$$
$$a'_{pp} = a_{pp}\cos^2\varphi - 2a_{pq}\sin\varphi\cos\varphi + a_{qq}\sin^2\varphi, \tag{3.7}$$
$$a'_{qq} = a_{pp}\sin^2\varphi + 2a_{pq}\sin\varphi\cos\varphi + a_{qq}\cos^2\varphi, \tag{3.8}$$

Iz (3.3) i (3.4) sijedi

$$0 = a_{pq}\underbrace{(\cos^2\varphi - \sin^2\varphi)}_{\cos 2\varphi} + (a_{pp} - a_{qq})\underbrace{\sin\varphi\cos\varphi}_{\frac{1}{2}\sin 2\varphi}. \tag{3.9}$$

Ekvivalentno,

$$\operatorname{ctg} 2\varphi = \frac{a_{qq} - a_{pp}}{2a_{pq}}. \tag{3.10}$$

Neka je $\zeta \equiv \operatorname{ctg} 2\varphi$, $\tau \equiv \operatorname{tg}\varphi$. Trigonometrijske formule daju $\zeta = (1 - \tau^2)/(2\tau)$, pa τ možemo dobiti kao rješenje kvadratne jednadžbe

$$\tau^2 + 2\zeta\tau - 1 = 0.$$

Rješenja su

$$\tau_{\pm} = \frac{-2\zeta \pm \sqrt{4\zeta^2 + 4}}{2} = -\zeta \pm \sqrt{\zeta^2 + 1}.$$

Zbog stabilnosti uzimamo po modulu manje rješenje koje možemo zapisati kao

$$\tau = \frac{\operatorname{sign}(\zeta)}{|\zeta| + \sqrt{\zeta^2 + 1}}. \tag{3.11}$$

To rješenje odgovara kutu rotacije φ za koji vrijedi $|\varphi| \leq \pi/4$. Iz trigonometrijskih formula slijedi

$$\cos\varphi = \frac{1}{\sqrt{1 + \tau^2}},\ \sin\varphi = \tau\cos\varphi = \frac{\tau}{\sqrt{1 + \tau^2}}. \tag{3.12}$$

Uz (3.9), izrazi (3.7) i (3.8) postaju

$$\begin{aligned} a'_{pp} &= a_{pp}\cos^2\varphi - 2a_{pq}\sin\varphi\cos\varphi + (\frac{\cos^2\varphi - \sin^2\varphi}{\sin\varphi\cos\varphi}a_{pq} + a_{pp})\sin^2\varphi \\ &= a_{pp}(\sin^2\varphi + \cos^2\varphi) - \frac{\sin\varphi\cos^2\varphi - \sin^3\varphi - 2\sin\varphi\cos^2\varphi}{\cos\varphi}\, a_{pq} \\ &= a_{pp} + \frac{\sin\varphi(-\sin^2\varphi - \cos^2\varphi)}{\cos\varphi}\, a_{pq} \\ &= a_{pp} - \tau a_{pq}, \end{aligned} \tag{3.13}$$

$$\begin{aligned} a'_{qq} &= (a_{qq} - \frac{\cos^2\varphi - \sin^2\varphi}{\sin\varphi\cos\varphi}a_{pq})\sin^2\varphi + 2a_{pq}\sin\varphi\cos\varphi + a_{qq}\cos^2\varphi \\ &= a_{qq}(\sin^2\varphi + \cos^2\varphi) + \frac{-\sin\varphi\cos^2\varphi + \sin^3\varphi + 2\sin\varphi\cos^2\varphi}{\cos\varphi}\, a_{pq} \\ &= a_{qq} + \frac{\sin\varphi(\sin^2\varphi + \cos^2\varphi)}{\cos\varphi}\, a_{pq} \\ &= a_{qq} + \tau a_{pq}. \end{aligned} \tag{3.14}$$

Imamo i

$$\begin{aligned} a'_{rp} &= \cos\varphi a_{rp} - \sin\varphi a_{rq} = a_{rp} + (\cos\varphi - 1)a_{rp} - \sin\varphi a_{rq} \\ &= a_{rp} - \sin\varphi\Big(a_{rq} + \frac{1-\cos\varphi}{\sin\varphi}a_{rp}\Big), \\ a'_{rq} &= \cos\varphi a_{rq} + \sin\varphi a_{rp} = a_{rq} + (\cos\varphi - 1)a_{rq} + \sin\varphi a_{rp} \\ &= a_{rq} + \sin\varphi\Big(a_{rp} - \frac{1-\cos\varphi}{\sin\varphi}a_{rq}\Big) \end{aligned}$$

gdje je

$$\frac{1-\cos\varphi}{\sin\varphi} = \frac{(1-\cos\varphi)(1+\cos\varphi)}{\sin\varphi(1+\cos\varphi)} = \frac{1-\cos^2\varphi}{\sin\varphi(1+\cos\varphi)} = \frac{\sin\varphi}{1+\cos\varphi} = \operatorname{tg}\frac{\varphi}{2}.$$

Sada dajemo rezultate o točnosti Jacobijeve rotacije. Dokazi tvrdnji su u Dodatku.

Lema 3.1 ([4])**.** *Neka je τ definiran s (3.11). Tada vrijedi*

$$fl(\tau) = \tau(1+\varepsilon_\tau), \quad |\varepsilon_\tau| \leq 7\varepsilon_M.$$

Teorem 3.2 ([4]). *Za c i s dane s (3.12) vrijedi*

$$fl(c) = c(1+\varepsilon_c), \quad |\varepsilon_c| \leq 10\varepsilon_M,$$
$$fl(s) = s(1+\varepsilon_s), \quad |\varepsilon_s| \leq 17\varepsilon_M.$$

Dakle, matrica $J_{(p,q)}$ iz (3.1) će biti točno izračunata. Rezultati o točnosti računanja matrice A' definirane s (3.2) se mogu naći u [4] i [5].

Matlab kod opisane rotacije je u Algoritmu 12.

Algoritam 12

```
function [A,Q] = Jacobi(A,p,q)
if A(p,q)<eps
    Q = eye(size(A)); return;
else
    t = (A(q,q)-A(p,p))/(2*A(p,q));
    t = sign(t)/(abs(t)+sqrt(1+t^2));
    c = 1/sqrt(t^2+1);
    s = t/sqrt(t^2+1);
    Q = eye(size(A)); Q(p,p) = c; Q(q,q) = c; Q(p,q) = s; Q(q,p) = -s;

    A(p,p) = A(p,p)-t*A(p,q);
    A(q,q) = A(q,q)+t*A(p,q);
    A(p,q) = 0; A(q,p) = 0;

    n = size(A,1); t = s/(1+c);
    for r = 1:n
    if r≠p && r≠q
           A(r,p) = A(r,p) - s*(A(r,q) + t*A(r,p));
           A(r,q) = A(r,q) + s*(A(p,r) - t*A(r,q));
           A(q,r) = A(r,q); A(p,r) = A(r,p);
    end
    end
end
end
```

Poglavlje 4

Nova metoda

U ovom poglavlju opisujemo metodu za spektralnu dekompoziciju simetričnih matrica reda 3 i 4 koja se zasniva na algoritmu *aheig* iz poglavlja 2. Taj algoritam kombiniramo s Jacobijevom rotacijom iz poglavlja 3 kako bi dobili spektralnu dekompoziciju simetrične matrice reda 3, a zatim i spektralnu dekompoziciju simetrične matrice reda 4.

4.1 Spektralna dekompozicija simetričnih matrica reda 3

Neka je

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

simetrična matrica reda 3. Ako je $a_{12} = 0$, onda je matrica A streličasta. U suprotnom te elemente možemo poništiti Jacobijevom rotacijom

$$J_{(1,2)} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

gdje je kut φ takav da vrijedi

$$\zeta = \operatorname{ctg} 2\varphi = \frac{a_{22} - a_{11}}{2a_{12}},\ \tau = \operatorname{tg}\varphi = \frac{\operatorname{sign}(\zeta)}{|\zeta| + \sqrt{\zeta^2 + 1}}. \tag{4.1}$$

Tada je

$$c \equiv \cos\varphi = \frac{1}{\sqrt{1+\tau^2}},\ s \equiv \sin\varphi = \tau\cos\varphi \tag{4.2}$$

i prema (3.3), (3.5), (3.6), (3.13), (3.14) vrijedi

$$B \equiv J_{(1,2)}^T A J_{(1,2)} = \begin{bmatrix} a_{11} - \tau a_{12} & 0 & a_{13}c - a_{23}s \\ 0 & a_{22} + \tau a_{12} & a_{23}c + a_{13}s \\ a_{13}c - a_{23}s & a_{23}c + a_{13}s & a_{33} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}.$$

Dobivena matrica B je streličasta. Imamo nekoliko mogućnosti.

- Ako je $b_{13} = b_{23} = 0$, onda je matrica B dijagonalna, a svojstveni vektori su stupci matrice $J_{(1,2)}$.
- Ako je $b_{13} = 0$, $b_{23} \neq 0$, onda je b_{11} svojstvena vrijednost matrice A i B se može dijagonalizirati Jacobijevom rotacijom u ravnini $(2, 3)$. Uz

  $$J_{(2,3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_1 & s_1 \\ 0 & -s_1 & c_1 \end{bmatrix}, \; s_1 = c_1 t, \; c_1 = \frac{1}{\sqrt{1+t^2}}, \; t = \frac{\operatorname{sign}(\xi)}{|\xi| + \sqrt{\xi^2 + 1}}, \; \xi = \frac{b_{33} - b_{22}}{2b_{23}},$$

  je

  $$C \equiv (JJ_{(2,3)})^T A (JJ_{(2,3)}) = \begin{bmatrix} a_{11} - \tau a_{12} & 0 & 0 \\ 0 & a_{22} + \tau a_{12} - t(a_{23}c + a_{13}s) & 0 \\ 0 & 0 & a_{33} + t(a_{23}c + a_{13}s) \end{bmatrix}$$

  dijagonalna matrica koja na dijagonali ima svojstvene vrijednosti matrice A. Svojstveni vektori od A su stupci (ortogonalne) matrice $J_{(1,2)}J_{(2,3)}$.
- Ako je $b_{23} = 0$, $b_{13} \neq 0$, onda je b_{22} svojstvena vrijednost matrice A i B se može dijagonalizirati Jacobijevom rotacijom u ravnini $(1, 3)$. Analogno prethodnom slučaju, uz

  $$J_{(1,3)} = \begin{bmatrix} c_1 & 0 & s_1 \\ 0 & 1 & 0 \\ -s_1 & 0 & c_1 \end{bmatrix}, \; s_1 = c_1 t, \; c_1 = \frac{1}{\sqrt{1+t^2}}, \; t = \frac{\operatorname{sign}(\xi)}{|\xi| + \sqrt{\xi^2 + 1}}, \; \xi = \frac{b_{33} - b_{11}}{2b_{13}},$$

  je

  $$C \equiv (J_{(1,2)}J_{(1,3)})^T A (J_{(1,2)}J_{(1,3)}) = \begin{bmatrix} a_{11} - \tau a_{12} - t(a_{13}c - a_{23}s) & 0 & 0 \\ 0 & a_{22} + \tau a_{12} & 0 \\ 0 & 0 & a_{33} + t(a_{13}c - a_{23}s) \end{bmatrix}$$

  dijagonalna matrica koja na dijagonali ima svojstvene vrijednosti matrice A. Svojstveni vektori od A su stupci (ortogonalne) matrice $J_{(1,2)}J_{(1,3)}$.

Do kraja odjeljka pretpostavljamo da vrijedi $b_{13} \neq 0, b_{23} \neq 0$.

- Ako vrijedi $b_{11} = b_{22} \equiv d$, onda koristimo Givensovu rotaciju

$$G_{(1,2)} = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Vrijedi

$$C \equiv G_{(1,2)}^T B G_{(1,2)} = \begin{bmatrix} d & 0 & \cos\psi b_{13} - \sin\psi b_{23} \\ 0 & d & \sin\psi b_{13} + \cos\psi b_{23} \\ \cos\psi b_{13} - \sin\psi b_{23} & \sin\psi b_{13} + \cos\psi b_{23} & b_{33} \end{bmatrix},$$

pa možemo odabrati ψ za koji je $\cos\psi b_{13} - \sin\psi b_{23} = 0$. Tada je

$$c_{23} = \sin\psi b_{13} + \cos\psi b_{23}, \quad \sin\psi = t \cdot \cos\psi, \quad \cos\psi = \frac{1}{\sqrt{1+t^2}}, \quad t \equiv \operatorname{tg}\psi = \frac{b_{13}}{b_{23}}$$

i matrica C se može dijagonalizirati Jacobijevom rotacijom u ravnini $(2,3)$. Uz

$$J_{(2,3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_1 & s_1 \\ 0 & -s_1 & c_1 \end{bmatrix}, \; s_1 = c_1 t, \; c_1 = \frac{1}{\sqrt{1+t^2}}, \; t = \frac{\operatorname{sign}(\xi)}{|\xi| + \sqrt{\xi^2+1}}, \; \xi = \frac{b_{33} - d}{2c_{23}},$$

je

$$D \equiv (J_{(1,2)} G_{(1,2)} J_{(2,3)})^T A (J_{(1,2)} G_{(1,2)} J_{(2,3)}) = \begin{bmatrix} d & 0 & 0 \\ 0 & d - tc_{23} & 0 \\ 0 & 0 & b_{33} + tc_{23} \end{bmatrix}$$

dijagonalna matrica koja na dijagonali ima svojstvene vrijednosti matrice A. Svojstveni vektori od A su stupci (ortogonalne) matrice $J_{(1,2)} G^T J_{(2,3)}$.

- Ako vrijedi $b_{11} < b_{22}$, onda uz permutaciju

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} (= P^{-1})$$

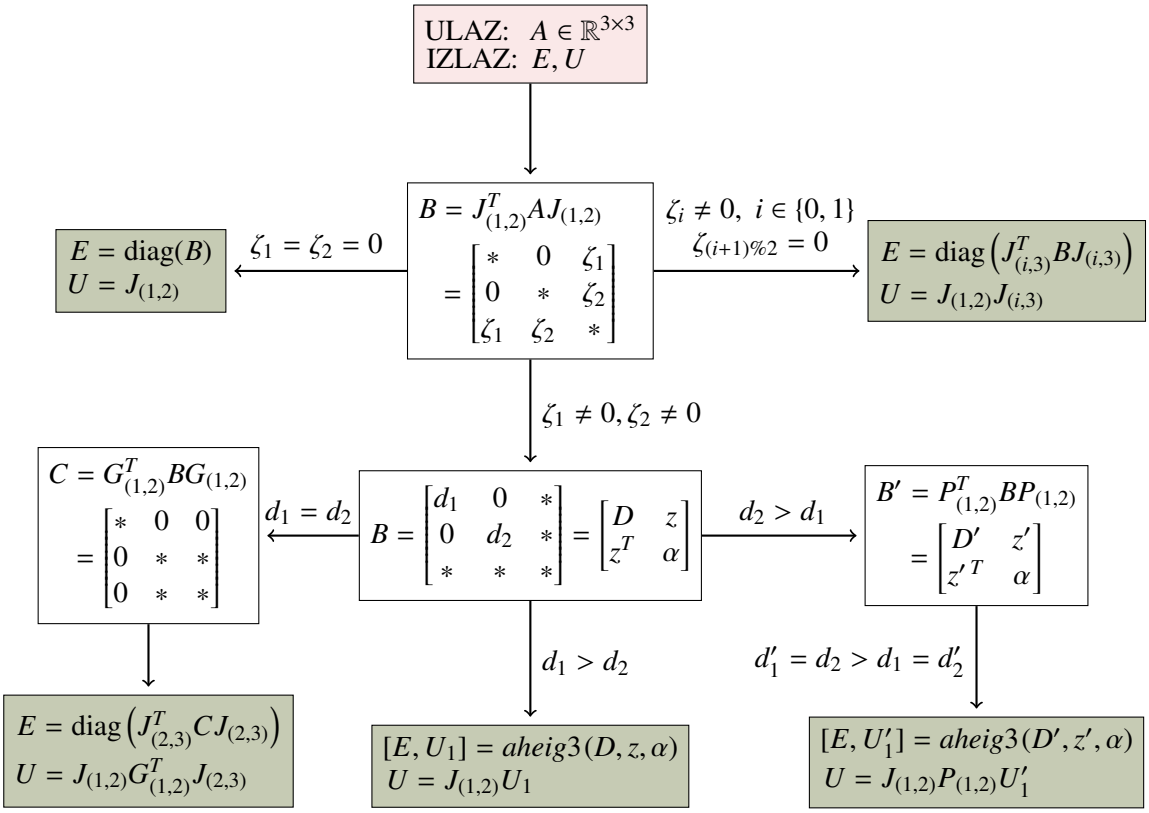
matrica

$$C = P^{-1} B P = \begin{bmatrix} b_{22} & 0 & b_{23} \\ 0 & b_{11} & b_{13} \\ b_{32} & b_{31} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{22} + \tau a_{12} & 0 & a_{23}c + a_{13}s \\ 0 & a_{11} - \tau a_{12} & a_{13}c - a_{23}s \\ a_{23}c + a_{13}s & a_{13}c - a_{23}s & a_{33} \end{bmatrix}$$

ima svojstvo $c_{11} > c_{22}$ i vrijedi

$$Cv = \lambda v \iff B(Pv) = \lambda(Pv).$$

Nadalje smatramo da vrijedi i $b_{11} > b_{22}$, pa primjenom algoritma *aheig3* (Algoritmi 8 i 9) dobivamo svojstvene vrijednosti $E = [\lambda_1, \lambda_2, \lambda_3]$ i svojstvene vektore matrice B, $U = [u_1\ u_2\ u_3]$. Tada su svojstveni vektori matrice A dani kao stupci matrice $J_{(1,2)}U$. Opisanu metodu zovemo *jaheig*3. Tok algoritma *jaheig*3 dajemo na Slici 4.1, a pripadni Matlab kod u Algoritmu 13. Matlab funkcije koje *jaheig*3 poziva su dane u Algoritmima 1, 12 i 8.



Slika 4.1: Tok algoritma *jaheig*3.

Algoritam 13

```
function [E,U] = jaheig3(A)
n = size(A,1);
[A,U] = Jacobi(A,1,2);

ctrl=-1;
if (abs(A(1,3))<eps && abs(A(2,3))<eps)
    ctrl=0;
elseif abs(A(1,3))<eps
    [A,Q] = Jacobi(A,2,3); ctrl=1;
elseif abs(A(2,3))<eps
    [A,Q] = Jacobi(A,1,3); ctrl=1;
elseif abs(A(1,1)-A(2,2))<eps
    [A,Q] = Givens(A,1,2); ctrl=1;
end
if ctrl≥0
    if n == 3
        E = diag(A);
    else
        E = A;
    end
    if ctrl==1
        U = U*Q;
    end
    return;
end

if A(1,1)>A(2,2)
    [E,Q] = aheig3([A(1,1); A(2,2)],A(1:2,3),A(3,3));
else
    [E,Q] = aheig3([A(2,2); A(1,1)],[A(2,3);A(1,3)],A(3,3));
    Q = [Q(2,:); Q(1,:); Q(3,:)];
end
if n == 4
    A(1,4) = A(4,1:3)*Q(:,1);
    A(2,4) = A(4,1:3)*Q(:,2);
    A(3,4) = A(4,1:3)*Q(:,3);
    A(4,1) = A(1,4); A(4,2) = A(2,4); A(4,3) = A(3,4);
    A(1:3,1:3) = diag(E);
    E = A;
    U1 = eye(4); U1(1:3,1:3) = Q;
    U = U*U1;
else
    U = U*Q;
end

end
```

Primjer 4.1. *Neka je*

$$A = \begin{bmatrix} 10^7 & 3 & 3 \\ 3 & 4 & 10^{-3} \\ 3 & 10^{-3} & 5 \end{bmatrix}.$$

Primjena Jacobijeve rotacije daje matricu

$$B = \begin{bmatrix} 1.0 \cdot 10^7 & 0.0 & 3.000000000299865 \\ 0.0 & 3.99999909999964 & 0.0009990999996399551 \\ 3.000000000299865 & 0.0009990999996399551 & 5.0 \end{bmatrix}.$$

*Primjenom algoritma jaheig*3 *na matricu B dobivamo svojstvene vrijednosti*

$$\begin{aligned} \lambda_1 &= 0.00020000015228986740, \\ \lambda_2 &= 10000000.000000000, \\ \lambda_3 &= 1000000000.000000, \end{aligned}$$

i svojstvene vektore kao stupce matrice U_{jaheig}*:*

$$\begin{bmatrix} -3.0030030030030103 \cdot 10^{-10} & -1.0 & 9.999997000200002 \cdot 10^{-11} \\ -1.0 & 3.003003003103103 \cdot 10^{-10} & 9.999999997000200 \cdot 10^{-11} \\ -1.0 \cdot 10^{-10} & -9.999996997196994 \cdot 10^{-11} & -1.0 \end{bmatrix}.$$

Uz $\Lambda_{jaheig} = [\lambda_1, \lambda_2, \lambda_3]$ *imamo*

$$\max |AU_{jaheig} - U_{jaheig} \operatorname{diag}(\Lambda_{jaheig})| = 1.332 \cdot 10^{-15}$$

i

$$\max |U^T_{jaheig3} U_{jaheig3} - I_3| = 6.661 \cdot 10^{-16}.$$

4.2 Spektralna dekompozicija simetričnih matrica reda 4

U ovom poglavlju opisujemo metodu za dijagonalizaciju simetričnih matrica reda 4 koja se zasniva na algoritmu *aheig* iz poglavlja 2. Taj algoritam kombiniramo s Jacobijevom rotacijom iz poglavlja 3 i algoritmom *jaheig*3 iz poglavlja 4.1 kako bi dobili spektralnu dekompoziciju simetrične matrice reda 4. Neka je

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}$$

simetrična matrica reda 4. Ako je $a_{12} = 0$, onda je podmatrica $A(1:3, 1:3)$ streličasta. U suprotnom te elemente možemo poništiti Jacobijevom rotacijom

$$J_{(1,2)} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 & 0 \\ -\sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

gdje je kut φ takav da vrijedi (4.1) i (4.2). Prema (3.3), (3.5), (3.6), (3.13), (3.14) imamo

$$B \equiv J_{(1,2)}^T A J_{(1,2)} = \begin{bmatrix} a_{11} - \tau a_{12} & 0 & a_{13}c - a_{23}s & a_{14}c - a_{24}s \\ 0 & a_{22} + \tau a_{12} & a_{23}c + a_{13}s & a_{24}c + a_{14}s \\ a_{13}c - a_{23}s & a_{23}c + a_{13}s & a_{33} & a_{34} \\ a_{14}c - a_{24}s & a_{24}c + a_{14}s & a_{34} & a_{44} \end{bmatrix},$$

pa je podmatrica $B(1:3, 1:3)$ streličasta, te postupamo kao u slučaju matrice reda 3.

- Ako je $b_{13} = b_{23} = 0$, onda je matrica B streličasta.
- Ako je $b_{13} = 0$, $b_{23} \neq 0$, onda Jacobijevom rotacijom u ravnini $(2, 3)$ dobivamo streličastu matricu.
- Ako je $b_{23} = 0$, $b_{13} \neq 0$, onda Jacobijevom rotacijom u ravnini $(1, 3)$ dobivamo streličastu matricu.

Neka je sada $b_{13} \neq 0, b_{23} \neq 0$.

- Ako vrijedi $b_{11} = b_{22} \equiv d$, onda pomoću (Givensove) rotacije

$$G_{(1,2)} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 & 0 \\ \sin\psi & \cos\psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

dobivamo matricu $C \equiv G_{(1,2)}^T B G_{(1,2)}$:

$$\begin{bmatrix} d & 0 & \cos\psi b_{13} - \sin\psi b_{23} & \cos\psi b_{14} - \sin\psi b_{24} \\ 0 & d & \sin\psi b_{13} + \cos\psi b_{23} & \sin\psi b_{14} + \cos\psi b_{24} \\ \cos\psi b_{13} - \sin\psi b_{23} & \sin\psi b_{13} + \cos\psi b_{23} & b_{33} & b_{34} \\ \cos\psi b_{14} - \sin\psi b_{24} & \sin\psi b_{14} + \cos\psi b_{24} & b_{34} & b_{44} \end{bmatrix}.$$

Možemo odabrati ψ za koji je $\cos\psi b_{13} - \sin\psi b_{23} = 0$. Tada je

$$c_{23} = \sin\psi b_{13} + \cos\psi b_{23}, \quad \sin\psi = t \cdot \cos\psi, \quad \cos\psi = \frac{1}{\sqrt{1+t^2}}, \quad t \equiv \operatorname{tg}\psi = \frac{b_{13}}{b_{23}},$$

te Jacobijevom rotacijom u ravnini $(2,3)$ dobivamo streličastu matricu.

- Ako vrijedi $b_{11} > b_{22}$, onda primjenom algoritma *aheig3* na matricu $B(1:3, 1:3)$ dobivamo njene svojstvene vrijednosti $\beta_1, \beta_2, \beta_3$ i svojstvene vektore $V = [v_1\ v_2\ v_3]$. Tada je

$$\begin{bmatrix} v_1^T & 0 & 0 \\ v_2^T & 0 & 0 \\ v_3^T & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} B \begin{bmatrix} & & & 0 \\ v_1 & v_2 & v_3 & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \beta_1 & 0 & 0 & B(4,1:3)v_1 \\ 0 & \beta_2 & 0 & B(4,1:3)v_2 \\ 0 & 0 & \beta_3 & B(4,1:3)v_3 \\ B(4,1:3)v_1 & B(4,1:3)v_2 & B(4,1:3)v_3 & b_{44} \end{bmatrix}$$

streličasta matrica.

- Ako vrijedi $b_{11} < b_{22}$, onda uz permutaciju

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} (= P^{-1})$$

matrica

$$C = P^{-1} B P = \begin{bmatrix} b_{22} & 0 & b_{23} & b_{24} \\ 0 & b_{11} & b_{13} & b_{14} \\ b_{32} & b_{31} & b_{33} & b_{34} \\ b_{42} & b_{41} & b_{43} & b_{44} \end{bmatrix}$$

ima svojstvo $c_{11} > c_{22}$ i vrijedi

$$Cv = \lambda v \iff B(Pv) = \lambda(Pv),$$

pa se ovaj slučaj može svesti na prethodni.

Trebamo još naći spektralnu dekompoziciju streličaste matrice reda 4. Označimo ju s

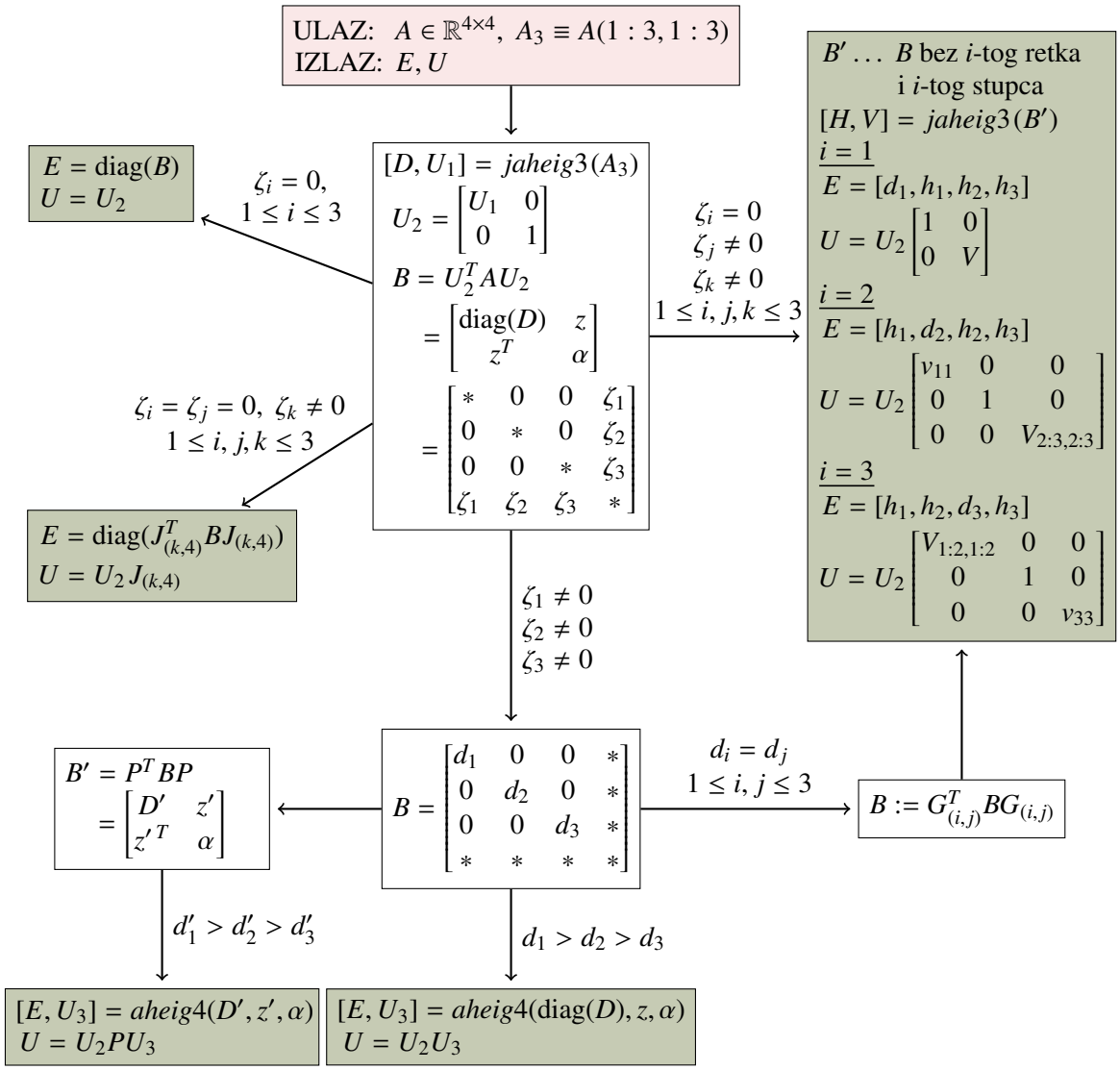
$$H = \begin{bmatrix} d_1 & 0 & 0 & \zeta_1 \\ 0 & d_2 & 0 & \zeta_2 \\ 0 & 0 & d_3 & \zeta_3 \\ \zeta_1 & \zeta_2 & \zeta_3 & \alpha \end{bmatrix}.$$

- Ako je $\zeta_i = 0$, za $i = 1, 2, 3$, matrica H je dijagonalna.
- Ako je $\zeta_i = \zeta_j = 0$, $\zeta_k \neq 0$, za međusobno različite $i, j, k \in \{1, 2, 3\}$, Jacobijevom rotacijom u ravnini $(k, 4)$ dobivamo dijagonalnu matricu.
- Ako je $\zeta_i = 0$, $\zeta_j \neq 0$, $\zeta_k \neq 0$, za međusobno različite $i, j, k \in \{1, 2, 3\}$, dovoljno je naći spektralnu dekompoziciju streličaste matrice reda 3 koju dobivamo kad iz A izbrišemo i-ti redak i i-ti stupac. Budući da ta matrica ne mora biti uređena, ne možemo koristiti algoritam *aheig*3, pa koristimo *jaheig*3 (pri čemu se Jacobijeva rotacija neće izvršiti jer je matrica već streličasta).

Neka je sada $\zeta_i \neq 0$, za $i = 1, 2, 3$.

- Ako je $d_i = d_j$, za neke $i, j \in \{1, 2, 3\}$, $i < j$, Givensovim rotacijama poništimo ζ_i i svedemo na prethodni slučaj.
- Ako ne vrijedi $d_1 > d_2 > d_3$, ispermutiramo retke i stupce.

Nadalje pretpostavljamo da je $d_1 > d_2 > d_3$, te primjenom algoritma *aheig4* (Algoritam 10) dobivamo svojstvene vrijednosti $E = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]$ i svojstvene vektore $U = [u_1\ u_2\ u_3\ u_4]$ matrice H. Opisanu metodu nazivamo *jaheig*4. Tok algoritma *jaheig*4 dajemo na Slici 4.2, a pripadni Matlab kod u Algoritmu 14. Matlab funkcije koje *jaheig*4 poziva su dane u Algoritmima 1, 12, 13 i 10.



Slika 4.2: Tok algoritma *jaheig*4.

Algoritam 14

```
function [E,U] = jaheig4(A)
[A,U] = jaheig3(A);
if (abs(A(1,4))<eps && abs(A(2,4))<eps && abs(A(3,4))<eps)
    E = diag(A);
elseif (abs(A(1,4))<eps && abs(A(2,4))<eps)
    [A,Q] = Jacobi(A,3,4); E = diag(A); U = U*Q;
elseif (abs(A(1,4))<eps && abs(A(3,4))<eps)
    [A,Q] = Jacobi(A,2,4); E = diag(A); U = U*Q;
elseif (abs(A(2,4))<eps && abs(A(3,4))<eps)
    [A,Q] = Jacobi(A,1,4); E = diag(A); U = U*Q;
elseif abs(A(1,4))<eps
    V = eye(4); [H,V(2:4,2:4)] = jaheig3(A(2:4,2:4));
    E = [A(1,1); H]; U = U*V;
elseif abs(A(2,4))<eps
    [H,V1] = jaheig3([A(1,1) A(1,3:4); A(3:4,1) A(3:4,3:4)]);
    E = [H(1); A(2,2); H(2:3)]; U = U*[V1(1,:); 0 1 0 0; V1(3:4,:)];
elseif abs(A(3,4))<eps
    [H,V1] = jaheig3([A(1:2,1) A(1:2,3:4); A(4,1) A(4,3:4)]);
    E = [H(1:2); A(3,3); H(3)]; U = U*[V1(1:2,:); 0 0 1 0; V1(4,:)];
elseif abs(A(1,1)-A(2,2))<eps
    [A,Q] = Givens(A,1,2); [H,V(2:4,2:4)] = jaheig3(A(2:4,2:4));
    E = [A(1,1); H]; U = U*Q*V;
elseif abs(A(1,1)-A(3,3))<eps
    [A,Q] = Givens(A,1,3); [H,V(2:4,2:4)] = jaheig3(A(2:4,2:4));
    E = [A(1,1); H]; U = U*Q*V;
elseif abs(A(2,2)-A(3,3))<eps
    [A,Q] = Givens(A,2,3); [H,V1] = jaheig3([A(1,1) A(1,3:4); ...
       A(3:4,1) A(3:4,3:4)]);
    E = [H(1); A(2,2); H(2:3)]; U = U*Q*[V1(1,:); 0 1 0 0; V1(3:4,:)];
elseif A(1,1)>A(2,2)
  if A(2,2)>A(3,3)
  [E,U1] = aheig4([A(1,1);A(2,2);A(3,3)],A(1:3,4),A(4,4));
  U = U*U1;
  elseif A(3,3)>A(1,1)
  [E,U1] = aheig4([A(3,3);A(1,1);A(2,2)],[A(3,4);A(1,4);A(2,4)],A(4,4));
  U = U*[U1(2:3,:);U1(1,:);U1(4,:)];
  else
  [E,U1] = aheig4([A(1,1);A(3,3);A(2,2)],[A(1,4);A(3,4);A(2,4)],A(4,4));
  U = U*[U1(1:2:3,:);U1(2,:);U1(4,:)];
  end
else
  if A(1,1)>A(3,3)
  [E,U1] = aheig4([A(2,2);A(1,1);A(3,3)],[A(2,4);A(1,4);A(3,4)],A(4,4));
  U = U*[U1(2,:);U1(1:2:3,:);U1(4,:)];
  elseif A(3,3)>A(2,2)
  [E,U1] = aheig4([A(3,3);A(2,2);A(1,1)],[A(3,4);A(2,4);A(1,4)],A(4,4));
  U = U*[U1(3:-1:1,:);U1(4,:)];
  else
  [E,U1] = aheig4([A(2,2);A(3,3);A(1,1)],[A(2,4);A(3,4);A(1,4)],A(4,4));
  U = U*[U1(3,:);U1(1:2,:);U1(4,:)];
  end
end
end
```

Dodatak

Ovdje dajemo dokaze tvrdnji iz odjeljka 2.2.

Veza točnosti svojstvenih vrijednosti matrica A i A_i

Dokaz teorema 2.5. Vrijedi

$$\widetilde{\lambda} \equiv fl(d_i + \widetilde{\mu}) = (d_i + \widetilde{\mu})(1 + \varepsilon_1), \quad |\varepsilon_1| \leq \varepsilon_M,$$

što uz (2.17, 2.18) postaje (zanemarujemo član reda $O(\varepsilon_M^2)$)

$$\begin{aligned} \lambda(1 + \kappa_\lambda \varepsilon_M) &= (d_i + \mu(1 + \kappa_\mu \varepsilon_M))(1 + \varepsilon_1) \\ \lambda + \lambda\kappa_\lambda \varepsilon_M &= d_i + d_i\varepsilon_1 + \mu + \mu(\kappa_\mu \varepsilon_M + \varepsilon_1). \end{aligned}$$

Vrijedi i $\lambda = \mu + d_i$, pa imamo

$$\lambda\kappa_\lambda \varepsilon_M = d_i\varepsilon_1 + \mu(\kappa_\mu \varepsilon_M + \varepsilon_1).$$

Uzimanjem apsolutnih vrijednosti konačno dobivamo

$$\begin{aligned} |\kappa_\lambda| = \Big|\frac{d_i\varepsilon_1 + \mu(\kappa_\mu \varepsilon_M + \varepsilon_1)}{\lambda \varepsilon_M}\Big| &\leq \frac{|d_i||\varepsilon_1| + \mu(|\kappa_\mu|\varepsilon_M + |\varepsilon_1|)}{|\lambda||\varepsilon_M|} \overset{|\varepsilon_1| \leq \varepsilon_M}{\leq} \frac{|d_i| + \mu(|\kappa_\mu| + 1)}{|\lambda|} \\ &\leq \frac{|d_i| + |\mu|}{|\lambda|}(|\kappa_\mu| + 1). \end{aligned}$$

□

Dokaz teorema 2.6.

(i) Zbog $\text{sign}(d_i) = \text{sign}(\mu)$ i $\lambda = d_i + \mu$ vrijedi

$$\frac{|d_i| + |\mu|}{|\lambda|} = \frac{|d_i + \mu|}{|d_i + \mu|} = 1.$$

(ii) Ako je $\text{sign}(d_i) = \text{sign}(\mu)$, tvrdnja slijedi iz (i). U suprotnom imamo

$$0 < d_{i+1} < \lambda < d_i, \quad \mu < 0,$$

ili

$$d_i < \lambda < d_{i-1} < 0, \quad \mu > 0.$$

U prvom slučaju je λ bliža polu d_i, pa vrijedi

$$|\mu| \leq \frac{1}{2}|d_i - d_{i+1}| \quad \text{i} \quad |\lambda| \geq \frac{1}{2}|d_i + d_{i+1}|.$$

Sada imamo

$$\frac{|d_i| + |\mu|}{|\lambda|} \leq \frac{|d_i| + \frac{1}{2}|d_i - d_{i+1}|}{\frac{1}{2}|d_i + d_{i+1}|} \overset{d_i > d_{i+1} > 0}{=} \frac{d_i + \frac{1}{2}d_i - \frac{1}{2}d_{i+1}}{\frac{1}{2}d_i + \frac{1}{2}d_{i+1}} \leq \frac{\frac{3}{2}d_i - \frac{1}{2}d_{i+1}}{\frac{1}{2}d_i + \frac{1}{2}d_{i+1}} \overset{d_{i+1} > 0}{\leq} \frac{3d_i}{d_i} = 3.$$

Drugi slučaj se dokazuje analogno.

$$\frac{|d_i| + |\mu|}{|\lambda|} \leq \frac{|d_i| + \frac{1}{2}|d_i - d_{i-1}|}{\frac{1}{2}|d_i + d_{i-1}|} = \frac{-d_i + \frac{1}{2}d_{i-.1} - \frac{1}{2}d_i}{-\frac{1}{2}d_i - \frac{1}{2}d_{i-1}} \leq \frac{-\frac{3}{2}d_i + \frac{1}{2}d_{i+1}}{-\frac{1}{2}d_i - \frac{1}{2}d_{i+1}} \leq \frac{-3d_i}{-d_i} = 3.$$

□

Dokaz korolara 2.7. Pretpostavimo prvo da nema promjene predznaka u elementima na dijagonali tj.

$$d_{n-1} < \cdots < d_1 < 0 \quad \text{ili} \quad 0 < d_{n-1} < \cdots < d_1.$$

U oba slučaja tvrdnja vrijedi za $\lambda_2, \ldots, \lambda_{n-1}$, po teoremu 2.6. U prvom slučaju je $\mu_n = \lambda_n - d_{n-1} < 0$, pa po teoremu 2.5 tvrdnja vrijedi i za λ_n. Ako λ_1 nije svojstvena vrijednost najbliža nuli, onda imamo

$$|\lambda_1| > |\lambda_2| > |d_1|,$$

pa je

$$\frac{|d_1| + |\lambda_1 - d_1|}{|\lambda_1|} \leq \frac{|d_1| + |\lambda_1| + |d_1|}{|\lambda_1|} < \frac{3|\lambda_1|}{|\lambda_1|} = 3.$$

Analogno, u drugom slučaju je $\mu_1 = \lambda_1 - d_1 > 0$, pa po teoremu 2.5 tvrdnja vrijedi i za λ_1. Ako λ_n nije svojstvena vrijednost najbliža nuli, onda je

$$|\lambda_n| > |\lambda_{n-1}| > |d_{n-1}|$$

i

$$\frac{|d_{n-1}| + |\lambda_n - d_{n-1}|}{|\lambda_n|} \leq \frac{|d_{n-1}| + |\lambda_n| + |d_{n-1}|}{|\lambda_n|} < \frac{3|\lambda_n|}{|\lambda_n|} = 3.$$

Pretpostavimo sada da postoji promjena predznaka. Postoji najviše jedna promjena predznaka jer je $d_1 > \cdots > d_{n-1}$ i neka je $d_j > 0 > d_{j+1}$. Tada po teoremu 2.5 tvrdnja vrijedi za λ_1 i λ_n, a za sve osim svojstvene vrijednosti λ_{j+1} tvrdnja vrijedi po teoremu 2.6. Ako λ_{j+1} nije najbliža nuli, onda je

$$|\lambda_{j+1}| > |\lambda_j| > |d_j| \quad \text{ili} \quad |\lambda_{j+1}| > |\lambda_{j+2}| > |d_{j+1}|.$$

Ako je u prvom slučaju najbliži pol d_j ili u drugom slučaju najbliži pol d_{j+1}, imamo

$$\frac{|d_j| + |\lambda_{j+1} - d_j|}{|\lambda_{j+1}|} \leq \frac{2|d_j| + |\lambda_{j+1}|}{|\lambda_{j+1}|} < 3 \quad \text{ili} \quad \frac{|d_{j+1}| + |\lambda_{j+1} - d_{j+1}|}{|\lambda_{j+1}|} \leq \frac{2|d_{j+1}| + |\lambda_{j+1}|}{|\lambda_{j+1}|} < 3.$$

Ako je u prvom slučaju najbliži pol d_{j+1}, onda je

$$\begin{aligned}\frac{|d_{j+1}| + |\lambda_{j+1} - d_{j+1}|}{|\lambda_{j+1}|} &\leq \frac{|d_{j+1} \pm \lambda_{j+1}| + |\lambda_{j+1} - d_{j+1}|}{|\lambda_{j+1}|} \leq \frac{2|\lambda_{j+1} - d_{j+1}| + |\lambda_{j+1}|}{|\lambda_{j+1}|} \\ &\leq \frac{2|\lambda_{j+1} - d_j| + |\lambda_{j+1}|}{|\lambda_{j+1}|} \leq \frac{3|\lambda_{j+1}| + 2|d_j|}{|\lambda_{j+1}|} < 5.\end{aligned}$$

Analogno, ako je u drugom slučaju najbliži pol d_j, onda je

$$\begin{aligned}\frac{|d_j| + |\lambda_{j+1} - d_j|}{|\lambda_{j+1}|} &\leq \frac{|d_j \pm \lambda_{j+1}| + |\lambda_{j+1} - d_j|}{|\lambda_{j+1}|} \leq \frac{2|\lambda_{j+1} - d_j| + |\lambda_{j+1}|}{|\lambda_{j+1}|} \\ &\leq \frac{2|\lambda_{j+1} - d_{j+1}| + |\lambda_{j+1}|}{|\lambda_{j+1}|} \leq \frac{3|\lambda_{j+1}| + 2|d_{j+1}|}{|\lambda_{j+1}|} < 5.\end{aligned}$$

□

Točnost svojstvenih vektora

Dokaz teorema 2.12. Neka su x i $\widetilde{x}$ definirani s (2.14) i (2.21), respektivno. Tvrdnja očito vrijedi za $x_n = \widetilde{x}_n = -1$. Za $\widetilde{x}_i$ imamo

$$\begin{aligned}x_i(1 + \varepsilon_{x_i}) = \widetilde{x}_i = fl(-\frac{\zeta_i}{\widetilde{\widetilde{\mu}}}) &= -\frac{\zeta_i}{\mu(1 + \kappa_\mu \varepsilon_M)}(1 + \varepsilon_1) = -\frac{\zeta_i}{\mu}\frac{(1 + \varepsilon_1)(1 - \kappa_\mu \varepsilon_M)}{1 - (\kappa_\mu \varepsilon_M)^2} \\ &= x_i \frac{1 + (\varepsilon_1 - \kappa_\mu \varepsilon_M) - \kappa_\mu \varepsilon_1 \varepsilon_M}{1 - (\kappa_\mu \varepsilon_M)^2}.\end{aligned}$$

Kad zanemarimo članove reda $O(\varepsilon_M^2)$ dobivamo

$$\widetilde{x_i} = x_i(1+\varepsilon_{x_i}), \quad |\varepsilon_{x_i}| = |\varepsilon_1 - \kappa_\mu \varepsilon_M| \leq (1+|\kappa_\mu|)\varepsilon_M.$$

Za $j \notin \{i, n\}$ rješavamo

$$x_j(1+\varepsilon_{x_j}) = \widetilde{x_j} = \frac{\zeta_j}{((d_j - d_i)(1+\varepsilon_2) - \mu(1+\kappa_\mu\varepsilon_M))(1+\varepsilon_3)}(1+\varepsilon_4)$$

po ε_{x_j}, pri čemu zanemarujemo članove reda $O(\varepsilon_M^2)$ i više. Dobivamo

$$\begin{aligned}
\frac{\zeta_j}{d_j-\lambda}\frac{1}{1-\varepsilon_{x_j}} &= \frac{\zeta_j}{(d_j-d_i)(1+\varepsilon_2+\varepsilon_3)-\mu(1+\kappa_\mu\varepsilon_M+\varepsilon_3)}\frac{1}{1-\varepsilon_4} \\
\frac{1}{d_j-\lambda}\frac{1}{1-\varepsilon_{x_j}} &= \frac{1}{(d_j-d_i)(1+\varepsilon_2+\varepsilon_3-\varepsilon_4)-\mu(1+\kappa_\mu\varepsilon_M+\varepsilon_3-\varepsilon_4)} \\
1-\varepsilon_{x_j} &= \frac{d_j-(d_i+\mu)+(d_j-d_i)(\varepsilon_2+\varepsilon_3-\varepsilon_4)-\mu(\kappa_\mu\varepsilon_M+\varepsilon_3-\varepsilon_4)}{d_j-\lambda} \\
\varepsilon_{x_j} &= \frac{(d_j-d_i)(-\varepsilon_2-\varepsilon_3+\varepsilon_4)+\mu(\kappa_\mu\varepsilon_M+\varepsilon_3-\varepsilon_4)}{d_j-\lambda},
\end{aligned}$$

pa je

$$\widetilde{x_j} = x_j(1+\varepsilon_{x_j}), \quad |\varepsilon_{x_j}| \leq \frac{|d_j-d_i|+|\mu|}{|d_j-\lambda|}(|\kappa_\mu|+3)\varepsilon_M.$$

Ako je $\operatorname{sign}(d_j - d_i) = -\operatorname{sign}(\mu)$, onda je

$$\frac{|d_j-d_i|+|\mu|}{|d_j-\lambda|} = \frac{|d_j-d_i-\mu|}{|d_j-\lambda|} = \frac{|d_j-\lambda|}{|d_j-\lambda|} = 1.$$

Ako je $\operatorname{sign}(d_j - d_i) = -\operatorname{sign}(\mu)$, onda, budući da je d_i najbliži pol vrijednosti λ, imamo

$$|\mu| \leq \frac{1}{2}|d_j - d_i|,$$

pa je

$$\frac{|d_j-d_i|+|\mu|}{|d_j-\lambda|} \leq \frac{|d_j-d_i-\mu|}{|d_j-d_i|-|\mu|} \leq \frac{\frac{3}{2}|d_j-d_i|}{\frac{1}{2}|d_j-d_i|} = 3.$$

Sve zajedno,

$$|\varepsilon_{x_k}| \leq 3(|\kappa_\mu|+3)\varepsilon_M, \quad k = 1, \ldots, n.$$

□

Točnost A_i^{-1}

Dokaz teorema 2.13. Za ne-nul elemente matrice A_i^{-1}, osim elementa b, imamo

$$fl\Big([A_i^{-1}]_{jj}\Big) = \frac{1}{(d_j - d_i)(1+\varepsilon_1)}(1+\varepsilon_2), \quad j \notin \{i, n\}, \tag{5.3}$$

$$fl\Big([A_i^{-1}]_{ji}\Big) = fl\Big([A_i^{-1}]_{ij}\Big) = \frac{-\zeta_j}{(d_j - d_i)(1+\varepsilon_3)\zeta_i(1+\varepsilon_4)}(1+\varepsilon_5), \quad j \notin \{i, n\}, \tag{5.4}$$

$$fl\Big([A_i^{-1}]_{ni}\Big) = fl\Big([A_i^{-1}]_{in}\Big) = \frac{1}{\zeta_i}(1+\varepsilon_6), \tag{5.5}$$

gdje je $|\varepsilon_k| \leq \varepsilon_M$, za $k = 1, \ldots, 6$. Opet zanemarujemo članove reda $O(\varepsilon_M^2)$ i više.

(5.3):

$$\begin{aligned}[A_i^{-1}]_{jj}(1+\varepsilon_{jj}) &= \frac{1}{(d_j - d_i)(1+\varepsilon_1)}(1+\varepsilon_2) = \frac{1}{(d_j - d_i)(1+\varepsilon_1-\varepsilon_2)} \\ &= [A_i^{-1}]_{jj}(1-\varepsilon_1+\varepsilon_2),\end{aligned}$$

pa imamo

$$fl\Big([A_i^{-1}]_{jj}\Big) = [A_i^{-1}]_{jj}(1+\varepsilon_{jj}), \quad |\varepsilon_{jj}| = |-\varepsilon_1+\varepsilon_2| \leq 2\varepsilon_M.$$

(5.4):

$$\begin{aligned}[A_i^{-1}]_{ij}(1+\varepsilon_{ij}) &= \frac{-\zeta_j}{(d_j - d_i)(1+\varepsilon_3)\zeta_i(1+\varepsilon_4)}(1+\varepsilon_5) \\ &= \frac{-\zeta_j}{(d_j - d_i)\zeta_i}(1-\varepsilon_3-\varepsilon_4+\varepsilon_5) \\ &= [A_i^{-1}]_{ij}(1-\varepsilon_3-\varepsilon_4+\varepsilon_5),\end{aligned}$$

pa imamo

$$fl\Big([A_i^{-1}]_{ij}\Big) = [A_i^{-1}]_{ij}(1+\varepsilon_{ij}), \quad |\varepsilon_{ij}| = |-\varepsilon_3-\varepsilon_4+\varepsilon_5| \leq 3\varepsilon_M.$$

(5.5): Odmah imamo

$$fl\Big([A_i^{-1}]_{in}\Big) = [A_i^{-1}]_{in}(1+\varepsilon_6), \quad |\varepsilon_6| \leq \varepsilon_M.$$

Sve zajedno, za sve elemente matrice A_i^{-1}, osim elementa $b = [A_i^{-1}]_{ii}$ vrijedi

$$fl\Big([A_i^{-1}]_{jk}\Big) = [A_i^{-1}]_{jk}(1+\varepsilon_{jk}), \quad |\varepsilon_{jk}| \leq 3\varepsilon_M.$$

□

Dokaz teorema 2.14.

(i) Za $k \in \{1, \ldots, n\}, k \neq i$ imamo

$$fl(\frac{\zeta_k^2}{d_k - d_i}) = \frac{\zeta_k^2(1+\varepsilon_1)}{(d_k - d_i)(1+\varepsilon_2)}(1+\varepsilon_3) = \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_1 - \varepsilon_2 + \varepsilon_3)$$
$$= \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_k), \quad |\varepsilon_k| \leq 3\varepsilon_M.$$

U sumi

$$\sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}$$

imamo $(i-1)$ član, pa se zbrajanje vrši $(i-2)$ puta. Svi članovi u sumi su istog predznaka, pa je

$$fl\left(\sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}\right) = \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_4), \quad |\varepsilon_4| \leq (i+1)\varepsilon_M.$$

Analogno,

$$fl\left(\sum_{k=i+1}^{n-1} \frac{\zeta_k^2}{d_k - d_i}\right) = \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_5), \quad |\varepsilon_5| \leq (n-i+1)\varepsilon_M.$$

Sada imamo

$$b + \delta b = \frac{1}{\zeta_i^2}\Bigg(- a(1+\varepsilon_a) + \sum_{k=1}^{i-1} \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_4) + \sum_{k=i+1}^{n-1} \frac{\zeta_k^2}{d_k - d_i}(1+\varepsilon_5)\Bigg)(1+\varepsilon_6),$$

uz $|\varepsilon_a| \leq \varepsilon_M$, $|\varepsilon_5| \leq 4\varepsilon_M$ pa je

$$|\delta b| \leq \frac{1}{\zeta_i^2}\Bigg(|-a| + \Big|\sum_{j=1}^{i-1} \frac{\zeta_j^2}{d_j - d_i}\Big| + \Big|\sum_{j=i+1}^{n-1} \frac{\zeta_j^2}{d_j - d_i}\Big|\Bigg)(n+5)\varepsilon_M,$$

tj.

$$|\delta b| \leq \frac{1}{\zeta_i^2}\Big(|a| + \left|z_1^T D_1^{-1} z_1\right| + \left|z_2^T D_2^{-1} z_2\right|\Big)(n+5)\varepsilon_M.$$

(ii) Iz relacija (2.13) i (2.19) i dijela (*i*) slijedi

$$|\kappa_b| = \frac{|\delta b|}{|b|}\frac{1}{\varepsilon_M} \leq \frac{\frac{1}{\zeta_i^2}(|a| + |z_1 D_1^{-1} z_1^T| + |z_2 D_2^{-1} Z_2^T|)(n+5)\varepsilon_M}{\frac{1}{\zeta_i^2}|-a + z_1 D_1^{-1} z_1^T + z_2 D_2^{-1} Z_2^T|}\frac{1}{\varepsilon_M}$$
$$\leq (n+5)\frac{|a| + |z_1 D_1^{-1} z_1^T| + |z_2 D_2^{-1} Z_2^T|}{|-a + z_1 D_1^{-1} z_1^T + z_2 D_2^{-1} Z_2^T|}.$$

□

Točnost bisekcije

Teorem 5.2 ([10])**.** *Neka je A streličasta matrica definirana u (2.1). Točnost računanja svojstvene vrijednosti $\lambda \in \sigma(A)$ bisekcijom je dana izrazom*

$$|\widetilde{\lambda} - \lambda| \leq \eta(\lambda),$$

gdje je

$$\eta(\lambda) = 1.06n\Big(|\alpha| + |\lambda| + \sum_{j=1}^{n-1} |\zeta_j|\Big)\varepsilon_M.$$

Dokaz. S $\widetilde{f_A}$ ćemo označiti funkciju f_A izračunatu s greškom zaokruživanja tj.

$$\widetilde{f_A}(x) = fl(f_A(x)),\ x \in \mathbb{R}.$$

Prvo dokazujemo pomoćnu tvrdnju.

Tvrdnja. Za $n > 2$, $\varepsilon_M < 0.001$ i $n\varepsilon_M < 0.1$ postoji simetrična matrica $H \in \mathbb{R}^{n\times n}$ takva da vrijedi

$$\widetilde{f_A}(\lambda) = f_{A+H}(\lambda)$$

i

$$|h_{nn}| \leq 1.06n(|\alpha| + |\lambda|)\varepsilon_M,$$
$$|h_{in}| = |h_{ni}| \leq 1.06n|\zeta_i|\varepsilon_M, \quad i = 1, \ldots, n-1,$$

a ostali elementi od H su nula.

Dokaz. Vidjeti [10]. □

Dakle, kada u aritmetici s pomičnim zarezom izračunamo $f_A(\lambda)$, dobijemo istu vrijednost koju bi dobili egzaktnim računom, ali s malo perturbiranom matricom $A + H$. Prema korolaru 1.5, takva perturbacija može pomaknuti svojstvene vrijednosti za najviše

$$|\lambda_k(A + H) - \lambda_k(A)| \leq \|H\|_2, \quad k = 1, \ldots, n. \tag{5.6}$$

Općenito za matricu $G \in \mathbb{R}^{n\times n}$ vrijedi

$$\|G\|_2 \leq \sqrt{\|G\|_1 \|G\|_\infty}.$$

Budući da je H simetrična, vrijedi

$$\|H\|_2 \leq \|H\|_\infty.$$

Iz tvrdnje i definicije norme $\|\cdot\|_\infty$ slijedi

$$\|H\|_\infty \leq 1.06n(|\alpha| + |\lambda| + \sum_{j=1}^{n-1} |\zeta_j|)\varepsilon_M. \tag{5.7}$$

Iz (5.6) i (5.7) slijedi tvrdnja. □

Lema 5.3 ([7, Korolar 4.7, Lema 4.1])**.** *Neka su A i λ_{max} definirane u (2.1) i (2.24), respektivno, te neka je $\lambda_k \in \sigma(A)$. Tada vrijedi*

$$\frac{|\widetilde{\lambda_k} - \lambda_k|}{|\lambda_{max}|} \leq 1.06n(\sqrt{n} + 1)\varepsilon_M.$$

Posebno,

$$\frac{|\widetilde{\lambda}_{max} - \lambda_{max}|}{|\lambda_{max}|} \leq 1.06n(\sqrt{n} + 1)\varepsilon_M.$$

Dokaz. Prema teoremu 5.2,

$$\begin{aligned}
\frac{|\widetilde{\lambda_k} - \lambda_k|}{|\lambda_{\max}|} &\leq \frac{1.06n}{|\lambda_{\max}|}\Big(|\alpha| + |\lambda_k| + \sum_{j=1}^{n-1} |\zeta_j|\Big)\varepsilon_M \\
&\leq \frac{1.06n}{|\lambda_{\max}|}\Big(\|A\|_\infty + |\lambda_k|\Big)\varepsilon_M \\
&\leq 1.06n\Big(\frac{\sqrt{n}\|A\|_2}{|\lambda_{\max}|} + \frac{|\lambda_k|}{|\lambda_{\max}|}\Big)\varepsilon_M \\
&\leq 1.06n(\sqrt{n} + 1)\varepsilon_M.
\end{aligned}$$

Pritom smo koristili

$$\left(|\alpha| + \sum_{j=1}^{n-1} |\zeta_j|\right) \leq \|A\|_\infty \leq \sqrt{n}\|A\|_2 = \sqrt{n}|\lambda_{\max}|, \quad |\lambda_k| \leq |\lambda_{\max}|.$$

□

Dokaz korolara 2.18. Iz

$$\widetilde{\lambda}_{\max} = \lambda_{\max}(1 + k_{\lambda_{\max}}\varepsilon_M)$$

slijedi

$$|k_{\lambda_{\max}}|\varepsilon_M \leq \frac{|\widetilde{\lambda}_{\max} - \lambda_{\max}|}{|\lambda_{\max}|}.$$

Iz leme 5.3 slijedi

$$|k_{\lambda_{\max}}|\varepsilon_M \leq \frac{|\widetilde{\lambda}_{\max} - \lambda_{\max}|}{|\lambda_{\max}|} \leq 1.06n(\sqrt{n} + 1)\varepsilon_M,$$

pa je

$$|k_{\lambda_{\max}}| \leq 1.06n(\sqrt{n} + 1).$$

□

Dokaz korolara 2.19. Iz

$$\widetilde{\lambda}_k = \lambda_k(1 + k_{\lambda_k}\varepsilon_M)$$

slijedi

$$|k_{\lambda_k}|\varepsilon_M \leq \frac{|\widetilde{\lambda}_k - \lambda_k|}{|\lambda_k|} = s \cdot \frac{|\widetilde{\lambda}_k - \lambda_k|}{|\lambda_{\max}|}.$$

Iz leme 5.3 slijedi

$$|k_{\lambda_k}|\varepsilon_M \leq s \cdot \frac{|\widetilde{\lambda}_k - \lambda_k|}{|\lambda_{\max}|} \leq s \cdot 1.06n(\sqrt{n} + 1)\varepsilon_M,$$

pa je

$$|k_{\lambda_k}| \leq s \cdot 1.06n(\sqrt{n} + 1).$$

□

Točnost svojstvenih vrijednosti matrice A_i

Lema 5.4 ([7, Teorem 5.3])**.** *Neka je A_i^{-1} matrica iz (2.12), ν_{max} njena po modulu najveća svojstvena vrijednost definirana u 2.25 te neka je*

$$\hat{\nu}_{max} = \nu_{max}(1 + k_{\nu_{max}}\varepsilon_M), \quad |k_{\nu_{max}}\varepsilon_M| < 1$$

odgovarajuća svojstvena vrijednost matrice $\widetilde{(A_i^{-1})}$. Tada vrijedi

$$|k_{\nu_{max}}| \leq 3\sqrt{n} + \frac{(n+5)}{|\nu_{max}|} \cdot \frac{|a| + |z_1^T D_1^{-1} z_1| + |z_2^T D_2^{-1} z_2|}{\zeta_i^2}.$$

Dokaz. Prema teoremu 2.13, možemo pisati

$$\widetilde{(A_i^{-1})} = A_i^{-1} + \delta A_i^{-1},$$

gdje je

$$\begin{aligned} \|\delta A_i^{-1}\|_2 &\leq \left\| \begin{bmatrix} |D_1^{-1}| & |w_1| & 0 & 0 \\ |w_1^T| & 0 & |w_2^T| & 1/|\zeta_i| \\ 0 & |w_2| & |D_2^{-1}| & 0 \\ 0 & 1/|\zeta_i| & 0 & 0 \end{bmatrix} \right\|_2 \cdot 3\varepsilon_M + \left\| \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & |\delta b| & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\|_2 \\ &\leq \| \, |A_i^{-1}| \, \|_2 \cdot 3\varepsilon_M + |\delta b| \leq \sqrt{n} \, \|A_i^{-1}\|_2 \cdot 3\varepsilon_M + |\delta b|. \end{aligned}$$

Prema teoremu 2.14 imamo

$$\|\delta A_i^{-1}\|_2 \leq \|A_i^{-1}\|_2 \cdot 3\varepsilon_M + \frac{1}{\zeta_i^2}(|a| + |z_1^T D_1 z_1| + |z_2^T D_2 z_2|)(n+5)\varepsilon_M.$$

$\hat{\nu}_{\max}$ i $\nu_{\max}$ su najveće svojstvene vrijednosti matrica $\widetilde{(A_i^{-1})}$ i A_i^{-1}, respektivno, pa vrijedi

$$|\hat{\nu}_{\max} - \nu_{\max}| = \|\widetilde{(A_i^{-1})}\|_2 - \|A_i^{-1}\|_2 \leq \|\delta A_i^{-1}\|_2.$$

Slijedi

$$|\nu_{\max} k_{\nu_{\max}}|\varepsilon_M \leq \|\delta A_i^{-1}\|_2 \leq \sqrt{n} \, \|A_i^{-1}\|_2 \cdot 3\varepsilon_M + \frac{(n+5)}{\zeta_i^2}(|a| + |z_1^T D_1 z_1| + |z_2^T D_2 z_2|)\varepsilon_M.$$

Konačno,

$$|k_{\nu_{\max}}| \leq 3\sqrt{n} + \frac{(n+5)}{|\nu_{\max}|} \frac{|a| + |z_1^T D_1 z_1| + |z_2^T D_2 z_2|}{\zeta_i^2}.$$

□

Lema 5.5 ([7, Teorem 5.4])**.** *Neka vrijede uvjeti teorema 5.4 i neka, uz oznake iz teorema 2.14, vrijedi*

$$|\kappa_b| \leq C.$$

Tada je

$$|k_{\nu_{max}}| \leq \sqrt{n}\max\{3, C\}.$$

Dokaz. Prema dokazu leme 5.4

$$\begin{aligned}\|\delta A_i^{-1}\|_2 &\leq \left\| \begin{bmatrix} |D_1^{-1}| & |w_1| & 0 & 0 \\ |w_1^T| & |b| & |w_2^T| & 1/|\zeta_i| \\ 0 & |w_2| & |D_2^{-1}| & 0 \\ 0 & 1/|\zeta_i| & 0 & 0 \end{bmatrix} \right\|_2 \cdot \max\{3, C\}\varepsilon_M \\ &\equiv \| \, |A_i^{-1}| \, \|_2 \cdot \max\{3, C\}\varepsilon_M \\ &\leq \sqrt{n} \, \| \, A_i^{-1} \, \|_2 \cdot \max\{3, C\}\varepsilon_M.\end{aligned}$$

Također, vrijedi

$$|\hat{\nu}_{\max} - \nu_{\max}| = \|\widetilde{(A_i^{-1})}\|_2 - \|A_i^{-1}\|_2 \leq \|\delta A_i^{-1}\|_2,$$

pa imamo

$$|\nu_{\max} k_{\nu_{\max}}|\varepsilon_M \leq \|A_i^{-1}\|_2 \leq \sqrt{n} \, \|A_i^{-1}\|_2 \cdot \max\{3, C\}\varepsilon_M.$$

Slijedi

$$|k_{\nu_{\max}}| \leq \sqrt{n}\max\{3, C\}.$$

□

Dokaz teorema 2.20. Slijedi direktno iz leme 5.4 i leme 5.5. □

Lema 5.6 ([7, Lema 5.1])**.** *Neka je A_i^{-1} definirana s (2.20). Tada za $k \in \{1, \ldots, n\}$, $k \neq i$ vrijedi*

$$\frac{|\zeta_k|}{|\zeta_i||d_k - d_i|} \leq \|A_i^{-1}\|_2.$$

Dokaz. Vrijedi

$$\|A_i^{-1}\|_2 \geq \|A_i^{-1}e_k\|_2 = \sqrt{\frac{1}{(d_k-d_i)^2} + \frac{\zeta_k^2}{\zeta_i^2(d_k-d_i)^2}} \geq \frac{|\zeta_k|}{|\zeta_i||d_k-d_i|}.$$

□

Dokaz teorema 2.21. Množeći matricu $\widetilde{A_i} = fl(A_i)$ dijagonalnom matricom

$$\begin{bmatrix} (1+\varepsilon_a)^{1/2} I_{n-1} & 0 \\ 0 & (1+\varepsilon_a)^{-1/2} \end{bmatrix}$$

s lijeva i s desna dobivamo

$$\begin{bmatrix} (1+\varepsilon_a)^{1/2} I_{i-1} & 0 & 0 & 0 \\ 0 & (1+\varepsilon_a)^{1/2} & 0 & 0 \\ 0 & 0 & (1+\varepsilon_a)^{1/2} I_{n-i-1} & 0 \\ 0 & 0 & 0 & (1+\varepsilon_a)^{-1/2} \end{bmatrix}.$$

$$\begin{bmatrix} D_1(I+E_1) & 0 & 0 & z_1 \\ 0 & 0 & 0 & \xi_i \\ 0 & 0 & D_2(I+E_2) & z_2 \\ z_1^T & \xi_i & z_2^T & a(1+\varepsilon_a) \end{bmatrix}.$$

$$\begin{bmatrix} (1+\varepsilon_a)^{1/2} I_{i-1} & 0 & 0 & 0 \\ 0 & (1+\varepsilon_a)^{1/2} & 0 & 0 \\ 0 & 0 & (1+\varepsilon_a)^{1/2} I_{n-i-1} & 0 \\ 0 & 0 & 0 & (1+\varepsilon_a)^{-1/2} \end{bmatrix}$$

$$= \begin{bmatrix} D_1(I+E_1)(1+\varepsilon_a) & 0 & 0 & z_1 \\ 0 & 0 & 0 & \xi_i \\ 0 & 0 & D_2(I+E_2)(1+\varepsilon_a) & z_2 \\ z_1^T & \xi_i & z_2^T & a \end{bmatrix}$$

Dakle, perturbacije u a možemo prikazati kao perturbacije u D_1 i D_2. Tada je

$$k_{\nu_{\max}} \leq 6\left(\sqrt{n} + \frac{1}{|\nu_{\max}|}\frac{|z_1^T D_1^{-1} z_1| + |z_2 D_2^{-1} z_2|}{\zeta_i^2}\right) \tag{5.8}$$

Sada korištenjem leme 5.6 i, dijeljenjem svakog člana $\frac{\zeta_k^2}{\zeta_i^2|d_k-d_i|}$ u (5.8) odgovarajućim $\frac{|\zeta_k|}{|\zeta_i||d_k-d_i|}$, dobivamo

$$k_{\nu_{\max}} \leq 6\left(\sqrt{n} + \frac{1}{\zeta_i}\sum_{\substack{k=1 \\ k\neq i}}^{n-1}|\zeta_k|\right) \leq 6\left(\sqrt{n} + (n-2)\frac{1}{\zeta_i}\max_{\substack{k=1,\ldots,n-1 \\ k\neq i}}|\zeta_k|\right).$$

□

Dokaz leme 2.22. Rješavamo

$$\begin{aligned}
\mu(1+k_\mu\varepsilon_M) &= fl\left(\frac{1}{\nu_{\max}(1+k_\nu\varepsilon_M)}\right)\\
\mu(1+k_\mu\varepsilon_M) &= \frac{1}{\nu_{\max}(1+k_\nu\varepsilon_M)}(1+\varepsilon_1)\\
\mu(1+k_\mu\varepsilon_M) &= \frac{1}{\nu_{\max}}(1-k_\nu\varepsilon_M+\varepsilon_1)\\
|k_\mu|\varepsilon_M &\le |k_\nu|\varepsilon_M+|\varepsilon_1| \le (|k_\nu|+1)\varepsilon_M\\
|k_\mu| &\le |k_\nu|+1.
\end{aligned}$$

□

Dokaz leme 3.1. Za veličinu ζ iz (3.10) imamo

$$\begin{aligned}
\zeta(1+\varepsilon_\zeta) = fl(\zeta) &= \frac{(a_{pp}-a_{qq})(1+\varepsilon_1)}{2a_{pq}(1+\varepsilon_2)}(1+\varepsilon_3) = \frac{a_{pp}-a_{qq}}{2a_{pq}}(1+\varepsilon_1-\varepsilon_2+\varepsilon_3)\\
&= \zeta(1+\varepsilon_1-\varepsilon_2+\varepsilon_3),
\end{aligned}$$

pa je

$$|\varepsilon_\zeta| \le |\varepsilon_1|+|\varepsilon_2|+|\varepsilon_3| \le 3\varepsilon_M.$$

Neka je sada

$$\tau_1 = \sqrt{\zeta^2+1}.$$

Računamo

$$\begin{aligned}
\tau_1(1+\varepsilon_{\tau_1}) &= \sqrt{((\zeta(1+\varepsilon_\zeta))^2(1+\varepsilon_4)+1)(1+\varepsilon_5)}(1+\varepsilon_6)\\
\tau_1^2(1+2\varepsilon_{\tau_1}) &= \zeta^2(2\varepsilon_\zeta+\varepsilon_4+\varepsilon_5+2\varepsilon_6)+(\varepsilon_5+2\varepsilon_6)\\
\tau_1^2 2|\varepsilon_{\tau_1}| &\le \zeta^2(1+2|\varepsilon_\zeta|+|\varepsilon_4|+|\varepsilon_5|+2|\varepsilon_6|)+(1+|\varepsilon_5|+2|\varepsilon_6|)\\
&\le (\zeta^2+1)10\varepsilon_M.
\end{aligned}$$

Dakle,

$$fl(\sqrt{\zeta^2+1}) = \sqrt{\zeta^2+1}(1+\varepsilon_{\tau_1}), \quad |\varepsilon_{\tau_1}| \le 5\varepsilon_M.$$

Konačno imamo

$$\tau(1+\varepsilon_\tau) = fl(\tau) = \frac{\operatorname{sign}(\zeta)}{(|\zeta| + \sqrt{\zeta^2+1}(1+\varepsilon_{\tau_1}))(1+\varepsilon_7)}(1+\varepsilon_8)$$
$$\frac{1-\varepsilon_\tau}{\tau} = \frac{1}{\operatorname{sign}(\zeta)}\Big(|\zeta|(1+\varepsilon_7-\varepsilon_8) + \sqrt{\zeta^2+1}(1+\varepsilon_{\tau_1}+\varepsilon_7-\varepsilon_8)\Big)$$
$$|\tau||\varepsilon_\tau| \le |\zeta|(|\varepsilon_7|+|\varepsilon_8|) + \sqrt{\zeta^2+1}(|\varepsilon_{\tau_1}|+|\varepsilon_7|+|\varepsilon_8|)$$
$$\le (|\zeta| + \sqrt{\zeta^2+1})7\varepsilon_M.$$

Dakle,

$$|\varepsilon_\tau| \le 7\varepsilon_M.$$

□

Dokaz teorema 3.2. Neka je

$$\tau_2 = \sqrt{\tau^2+1}.$$

Računamo

$$\tau_2(1+\varepsilon_{\tau_2}) = \sqrt{((\tau(1+\varepsilon_\tau))^2(1+\varepsilon_1)+1)(1+\varepsilon_2)}(1+\varepsilon_3)$$
$$\tau_1^2(1+2\varepsilon_{\tau_2}) = \tau^2(2\varepsilon_\tau+\varepsilon_1+\varepsilon_2+2\varepsilon_3) + (\varepsilon_2+2\varepsilon_3)$$
$$\tau_1^2 2|\varepsilon_{\tau_2}| \le \tau^2(1+2|\varepsilon_\tau|+|\varepsilon_1|+|\varepsilon_2|+2|\varepsilon_3|) + (1+|\varepsilon_2|+2|\varepsilon_3|)$$
$$\le (\tau^2+1)18\varepsilon_M$$

gdje smo koristili tvrdnju leme 3.1: $|\varepsilon_\tau| \le 7\varepsilon_M$. Dakle,

$$fl(\sqrt{\tau^2+1}) = \sqrt{\tau^2+1}(1+\varepsilon_{\tau_2}), \quad |\varepsilon_{\tau_2}| \le 9\varepsilon_M.$$

Sada uz (3.12) dobivamo

$$c(1+\varepsilon_c) = fl(c) = \frac{1}{\sqrt{\tau^2+1}(1+\varepsilon_{\tau_2})}(1+\varepsilon_4) = c(1-\varepsilon_{\tau_2}+\varepsilon_4),$$

pa onda i

$$|\varepsilon_c| \le |\varepsilon_{\tau_2}| + |\varepsilon_4| \le 10\varepsilon_M. \tag{5.9}$$

Iz (3.12) imamo

$$s(1+\varepsilon_s) = fl(s) = \frac{\tau(1+\varepsilon_\tau)}{\sqrt{\tau^2+1}(1+\varepsilon_{\tau_2})}(1+\varepsilon_5) = s(1+\varepsilon_\tau-\varepsilon_{\tau_2}+\varepsilon_5),$$

pa korištenjem (5.9) i leme 3.1 dobivamo

$$|\varepsilon_s| \le |\varepsilon_\tau| + |\varepsilon_{\tau_2}| + |\varepsilon_5| \le 17\varepsilon_M.$$

□

Bibliografija

[1] Golub G. H., *Some modified matrix eigenvalue problems*, SIAM Rev. 15, 318-334 (1973).

[2] Golub G. H. i Van Loan C. F., *Matrix computations*, Johns Hopkins University Press, 1996.

[3] Wilkinson J. H., *The algebraic eigenvalue problem*, Oxford University Press, 1988.

[4] Demmel J. i Veselić K., *Jacobi's method is more accurate than QR*.

[5] Matejaš J., *Accuracy of the Jacobi method on scaled diagonally dominant hermitian matrices*.

[6] Gu M. i Eisenstat S. C., *A Divide-and-Conquer Algorithm for the Symmetric Tridiagonal Eigenproblem*, SIAM J. Matrix Anal. Appl. 16.1, Jan. 1995, pp.172-191.

[7] Jakovčević Stor N., *Točan rastav svojstvenih vrijednosti streličastih matrica i primjene*, Doktorska disertacija, Sveučilište u Zagrebu, 2011.

[8] Jakovčević Stor N., Slapničar I., i Barlow J. L., *Accurate eigenvalue decomposition of arrowhead matrices and applications*, Linear Algebra Appl, 2015, DOI:10.1016/j.laa.2013.10.007.

[9] Parlett B. N., *The symmetric eigenvalue problem*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[10] O'Leary D. P. i Stewart G. W., *Computing the eigenvalues and eigenvectors of symmetric arrowhead matrices*, J. Comput. Phys. 90.2, 1990, pp.497-505.

[11] Drmač Z., *Numerička matematika*, 2010.

Sažetak

U ovom radu prezentiramo novu metodu za spektralnu dekompoziciju simetričnih matrica reda 3 i 4, koja se bazira na algoritmu *aheig* (ArrowHEead EIGenvalues/eigenvectors) za spektralnu dekompoziciju streličastih matrica.

U poglavlju 1 uvodimo osnovne pojmove i rezultate koje kasnije koristimo. Zatim detaljno opisujemo i analiziramo algoritam *aheig* (poglavlje 2), te ga kombiniramo s Jacobijevim rotacijama (poglavlje 3) kako bi dobili algoritam za spektralnu dekompoziciju simetričnih matrica reda 3 i 4 (poglavlje 4).

Prilažemo Matlab kodove svih opisanih algoritama, te dajemo odgovarajuće primjere.

Summary

We present a new method for the spectral decomposition of symmetric matrices of order 3 and 4, which is based on the algorithm *aheig* (ArrowHEead EIGenvalues/eigenvectors) for the spectral decomposition of arrowhead matrices.

In Chapter 1 we give some basic definitions and results. Then we describe and analyze the *aheig* algorithm (Chapter 2) and combine it with Jacobi rotations (Chapter 3) to get an algorithm for the spectral decomposition of symmetric matrices of order 3 and 4 (Chapter 4).

We implement all the presented algorithms in Matlab and show some interesting examples.

Životopis

Marija Kranjčević je rođena je 10. svibnja 1992. u Zagrebu, gdje je završila osnovnu i srednju školu. Tijekom osnovnog i srednjeg obrazovanja sudjelovala je na mnogim natjecanjima iz matematike i fizike, ostvarujući značajne uspjehe među kojima su dva prva mjesta na Državnom natjecanju iz fizike (2006. i 2008.), te brončana medalja na Međunarodnoj prirodoslovnoj olimpijadi mladih (*International Junior Science Olympiad*, Taiwan 2007.) Nakon završteka XV. gimnazije 2010. godine, upisala je Preddiplomski sveučilišni studij Matematika na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu, a zatim 2013. godine i Diplomski sveučilišni studij Primijenjena matematika. Za vrijeme studija je sudjelovala na tri ljetne škole (*International Summer School in Mathematics for Young Students*, Lyon 2012, *Design and security of cryptographic algorithms and devices for real-world applications*, Šibenik 2014. i *4th Lisbon Machine Learning School*, Lisabon 2014.) te provela deset tjedana radeći na projektu pod naslovom Towards a hybrid parallelization of Chebyshev Filtered Subspace Iteration u sklopu *Jülich Supercomputing Centre Guest student programme on Scientific Computing*, Jülich 2014. S rezultatom tog projekta je sudjelovala na konferenciji *SIAM Computational Science and Engineering*, Salt Lake City 2015. kao jedna od osam finalista natjecanja *5th BGCE Student Paper Prize*.