

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Pap Sándor

**NUMERIKUS MÓDSZEREK MÁTRIXOK
SAJÁTÉRTÉKEINEK SZÁMÍTÁSÁRA**

BSc Szakdolgozat

Témavezető:

Lócsi Levente, tanársegéd

Numerikus Analízis Tanszék



Budapest, 2014

Köszönetnyilvánítás

Köszönettel tartozom témavezetőmnek, Lócsi Leventének, a sok ötletért és tanácsért, valamint, hogy bármikor fordulhattam hozzá segítségért. Továbbá köszönöm családom és barátaim támogatását.

Tartalomjegyzék

1. Alapismeretek és motiváció	5
1.1. Definíciók és alapvető tételek	5
1.1.1. Mátrixok sajátértékei és sajátvektorai	5
1.1.2. Diagonalizálhatóság	8
1.1.3. Sajátérték-feladatok kondicionáltsága	11
1.2. Motiváció	13
1.2.1. Stabilitási problémák	13
1.2.2. Internetes oldalak rangsorolása	15
2. A sajátértékeket egyenként közelítő eljárások	16
2.1. A hatványmódszer	17
2.1.1. Konvergenciafeltételek	18
2.1.2. A hatványmódszer algoritmus	22
2.1.3. Az eltolás	23
2.2. Inverz iteráció	24
2.2.1. Konvergenciafeltételek	24
2.3. Householder-eljárás	26
3. A sajátértékeket egyszerre közelítő eljárások	29
3.1. A Jacobi-módszer	29
3.1.1. Leírás és konvergencia	29
3.1.2. Gyakorlati szempontok	31
3.2. QR -módszer	33
4. Megvalósítás MATLAB-ban, tesztfeladatok	37
4.1. Módszerek kódja	38

4.2. Tesztfeladatok	40
Irodalomjegyzék	42

1. fejezet

Alapismeretek és motiváció

Az első fejezetben bevezetek néhány fontosabb definíciót és tételt, melyek később szükségessé lesznek, majd mutatok kettő motivációt. Ebből az egyik fizikai példa lesz, egy pedig még a matematikával és fizikával kevésbé foglalkozó emberek számára is érdekes lehet. Ez az internetes oldalak rangsorolásának elvét mutatja be. Manapság mindenki használ internetes keresőket, mégis azt gondolom, kevesen tudják, hogyan is működnek ezek.

Az definíciók és alaptételek című alfejezet [4], míg a motiváció című [1] alapján készült.

1.1. Definíciók és alapvető tételek

1.1.1. Mátrixok sajátértékei és sajátvektorai

1.1.1. Definíció. Legyen $\mathbf{A} \in \mathbb{C}^{n \times n}$ egy tetszőleges négyzetes mátrix. Ha egy $\mathbf{0} \neq \mathbf{v} \in \mathbb{C}^n$ vektor és egy $\lambda \in \mathbb{C}$ szám esetén teljesül az

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

egyenlőség, akkor a \mathbf{v} vektort a mátrix sajátvektorának, a λ számot a sajátvektorhoz tartozó sajátértéknek nevezzük. Egy összetartozó sajátértéket és sajátvektort sajátpárnak nevezzük.

1.1.2. Tétel. Egy mátrix adott sajátértékhez tartozó sajátvektorai a nullvektorral kiegészítve \mathbb{C}^n egy alterét alkotják.

Bizonyítás. Elég azt belátni, hogy ha \mathbf{v}_1 és \mathbf{v}_2 két különböző, λ -hoz tartozó sajátvektor, akkor c_1, c_2 számok esetén $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \neq \mathbf{0}$ is sajátvektor. Ez pedig következik az

$$\mathbf{A}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{A}\mathbf{v}_1 + c_2\mathbf{A}\mathbf{v}_2 = c_1\lambda\mathbf{v}_1 + c_2\lambda\mathbf{v}_2 = \lambda(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)$$

egyenlőségből. \square

1.1.3. Tétel. *Egy $\mathbf{A} \in \mathbb{C}^{n \times n}$ mátrixnak a multiplicitást is figyelembe véve pontosan n darab sajátértéke van. A sajátértékek a $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ egyenlet megoldásai. Egy adott λ sajátértékhez tartozó \mathbf{v} sajátvektorokat az $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ lineáris algebrai egyenletrendszer nullvektortól különböző megoldásai adják.*

Bizonyítás. Az $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ egyenlőséget átrendezve kapjuk, hogy a sajátvektornak az $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ egyenletrendszer nullvektortól különböző megoldásának kell lennie. Ez az egyenletrendszer homogén, így biztosan van megoldása, mivel a nullvektor megoldás lesz. Ahhoz, hogy λ sajátérték legyen, pontosan az kell, hogy legyen nullvektortól különböző megoldása is az egyenletrendszernek. Ez pontosan akkor teljesül, ha $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Az egyenletből látszik, hogy a sajátértékek a $p_{\mathbf{v}}(\lambda) := \det(\mathbf{A} - \lambda\mathbf{I})$ polinom zérushelyei. Ezekről pedig az algebra alaptételéből tudjuk, hogy multiplicitással együtt n darab van belőlük. \square

1.1.4. Definíció. *Egy adott $\mathbf{A} \in \mathbb{C}^{n \times n}$ mátrix esetén a $p_{\mathbf{v}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ polinomot a mátrix karakterisztikus polinomjának, a $P_{\mathbf{A}}(\lambda) = 0$ egyenletet pedig karakterisztikus egyenletnek nevezzük.*

1.1.5. Tétel. *Jelölje $\lambda_1, \lambda_2, \dots, \lambda_n$ az $\mathbf{A} \in \mathbb{C}^{n \times n}$ mátrix sajátértékeit. Ekkor*

$$\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i, \quad \text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i,$$

ahol $\text{tr}(\mathbf{A})$ a mátrix nyoma, azaz a főátlóban szereplő elemek összege.

Bizonyítás. Tudjuk, hogy $p(\lambda) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ n -edfokú polinom $\lambda_1, \dots, \lambda_n$ zérushelyeire igazak az alábbi formulák:

$$\lambda_1\lambda_2 \cdots \lambda_n = (-1)^n \frac{a_0}{a_n}, \quad \lambda_1 + \lambda_2 + \dots + \lambda_n = -\frac{a_{n-1}}{a_n}.$$

A karakterisztikus polinom legmagasabb fokú tagjának együtthatója $a_n = (-1)^n$,

$$a_{n-1} = (-1)^{n-1}(a_{11} + \dots + a_{nn}) = (-1)^{n-1}(\text{tr}(\mathbf{A}))$$

és a szabad tag $a_0 = \det(\mathbf{A})$. Ezekből pedig következik az állítás. \square

1.1.6. Tétel. Egy $\mathbf{A} \in \mathbb{C}^{n \times n}$ mátrixnak pontosan akkor nincs nulla sajátértéke, ha nonszinguláris, azaz ha van inverze.

Bizonyítás. Tekintsük az $\mathbf{A}\mathbf{v} = \mathbf{0}$ homogén lineáris egyenletrendszerét. Ennek az egyenletnek pontosan akkor a nullvektor az egyetlen megoldása, ha $\det(\mathbf{A}) = \det(\mathbf{A} - 0\mathbf{I}) \neq 0$. Ez egyenértékű azzal, hogy a nulla nem sajátértéke a mátrixnak. \square

1.1.7. Definíció. Egy $\mathbf{A} \in \mathbb{C}^{n \times n}$ mátrix legnagyobb abszolút értékű sajátértékének abszolút értékét \mathbf{A} spektrálsugarának hívjuk. Jelölés: $\rho(\mathbf{A})$.

A sajátértékek komplex számsíkon való elhelyezkedéséről ad becslést az alábbi ún. Gersgorin-tétel.

1.1.8. Tétel. (Gersgorin-tétel) Tekintsük az $\mathbf{A} \in \mathbb{C}^{n \times n}$ mátrixot. Legyen \mathbf{K}_i a komplex számsíkon az a zárt körlap, melynek középpontja a_{ii} , és sugara $\sum_{j=1, j \neq i}^n |a_{ij}|$ ($i = 1, \dots, n$). Ekkor a mátrix sajátértékei az $\cup_{i=1, \dots, n} \mathbf{K}_i$ halmazban találhatóak.

Bizonyítás. Legyen λ egy sajátértéke a mátrixnak. Ha λ megegyezik valamelyik diagonális elemmel, akkor erre a sajátértékre igaz az állítás. Különben írjuk fel \mathbf{A} -t $\mathbf{A} = \mathbf{D} + \mathbf{T}$ alakban, ahol $\mathbf{D} = \text{diag}(\text{diag}(\mathbf{A}))$ az \mathbf{A} diagonálisát tartalmazó mátrix. Az $\mathbf{A} - \lambda\mathbf{I}$ mátrix szinguláris, így van olyan $\mathbf{x} \neq \mathbf{0}$ vektor, mellyel $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$, azaz $(\mathbf{D} - \lambda\mathbf{I})\mathbf{x} = -\mathbf{T}\mathbf{x}$. A bal oldali mátrix invertálható, hiszen olyan diagonális mátrix, melynek egyik főátlóbeli eleme sem nulla. Így

$$\|\mathbf{x}\|_\infty \leq \|(\mathbf{D} - \lambda\mathbf{E})^{-1}\mathbf{T}\|_\infty \|\mathbf{x}\|_\infty,$$

amiből \mathbf{x} -val való osztással kapjuk, hogy

$$1 \leq \frac{\sum_{j=1, j \neq k}^n |a_{kj}|}{|a_{kk} - \lambda|}$$

valamely $k = 1, \dots, n$ indexre, azaz λ a \mathbf{K}_k körlap belsejébe esik. \square

1.1.9. Megjegyzés. Ha s darab körlap uniója diszjunkt a többi körlaptól, akkor az unióban pontosan s darab sajátérték van. Ez az ún. második Gersgorin-tétel.

1.1.10. Tétel. Különböző sajátértékekhez tartozó sajátvektorok lineárisan függetlenek.

Bizonyítás. Elég a tételt csak két sajátértékre igazolni úgy, hogy megmutatjuk, hogy az egyikhez tartozó egy sajátvektor nem fejezhető ki a másikhoz tartozó sajátvektorok lineáris kombinációjaként. Indirekt tegyük fel, hogy egy \mathbf{A} mátrixnak $\lambda \neq \mu$ két sajátértéke, továbbá, hogy $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ és $\mathbf{A}\mathbf{w}_i = \mu\mathbf{w}_i$ ($i = 1, \dots, l$) esetén a \mathbf{v} sajátvektor $\mathbf{v} = \sum_{i=1}^l \alpha_i \mathbf{w}_i$ alakban írható megfelelő α_i konstansokkal. Ekkor

$$\lambda\mathbf{v} = \mathbf{A}\mathbf{v} = \mathbf{A} \sum_{i=1}^l \alpha_i \mathbf{w}_i = \mu \sum_{i=1}^l \alpha_i \mathbf{w}_i = \mu\mathbf{v},$$

ami csak úgy lehetne, ha $\lambda = \mu$. Ez ellentmondás, így az állítás igaz. \square

1.1.11. Következmény. Ha egy $n \times n$ -es mátrixnak minden sajátértéke különböző, akkor van n darab lineárisan független sajátvektorrendszer, azaz választható a sajátvektorai közül n darab lineárisan független vektor.

1.1.2. Diagonalizálhatóság

1.1.12. Definíció. Az \mathbf{A} és \mathbf{B} ugyanolyan méretű négyzetes mátrixokat hasonlónak hívjuk, ha van olyan \mathbf{S} reguláris mátrix, melyre $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$.

A hasonlóság ekvivalenciareláció.

1.1.13. Tétel. Hasonló mátrixok sajátértékei megegyeznek.

Bizonyítás. Azt kell belátni, hogy a két hasonló \mathbf{A} és \mathbf{B} mátrix karakterisztikus polinomja ugyanaz. Legyen $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$. Mivel $\det(\mathbf{S}) \cdot \det(\mathbf{S}^{-1}) = \det(\mathbf{I}) = 1$, ezért

$$\det(\mathbf{B} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}) \cdot \det(\mathbf{A} - \lambda\mathbf{I}) \cdot \det(\mathbf{S}) = \det(\mathbf{A} - \lambda\mathbf{I}).$$

\square

1.1.14. Megjegyzés. Hasonló mátrixok sajátvektorai között fennáll az alábbi összefüggés: ha \mathbf{v} sajátvektora \mathbf{B} -nek, akkor $\mathbf{S}\mathbf{v}$ sajátvektora \mathbf{A} -nak.

1.1.15. Definíció. Egy \mathbf{A} mátrixot diagonalizálhatónak hívunk, ha hasonló egy diagonális mátrixhoz.

1.1.16. Tétel. Egy $n \times n$ -es mátrix pontosan akkor diagonalizálható, ha van n elemű lineárisan független sajátvektorrendszere.

Bizonyítás. Tegyük fel először, hogy az $n \times n$ -es \mathbf{A} mátrixnak van n darab lineárisan független sajátvektora, azaz $\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_j$ ($j = 1, \dots, n$) úgy, hogy a \mathbf{v}_j sajátvektorok lineárisan függetlenek. Ekkor igaz az

$$\mathbf{A} \underbrace{[\mathbf{v}_1 \dots \mathbf{v}_n]}_{:=\mathbf{S}} = \underbrace{[\mathbf{v}_1 \dots \mathbf{v}_n]}_{:=\mathbf{A}} \begin{bmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \dots \\ & & \ddots & \\ 0 & & & \dots \end{bmatrix}$$

egyenlőség. Legyen az $\mathbf{S} = [\mathbf{v}_1 \dots \mathbf{v}_n]$, ahol az \mathbf{S} oszlopvektorai sajátvektorok. Ekkor $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{\Lambda}$, azaz a mátrix diagonalizálható.

A másik irányhoz tegyük fel, hogy van olyan reguláris \mathbf{S} mátrix, mellyel $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{\Lambda}$, valamilyen $\mathbf{\Lambda}$ diagonális mátrixszal. Ekkor \mathbf{A} sajátértékei megegyeznek $\mathbf{\Lambda}$ elemeivel. Mivel az \mathbf{e}_j rendszer sajátvektorrendszere a $\mathbf{\Lambda}$ mátrixnak, így $\mathbf{S}\mathbf{e}_j$ sajátvektorrendszere \mathbf{A} -nak. Ezek \mathbf{S} regularitása miatt lineárisan független vektorok. \square

1.1.17. Definíció. Egy \mathbf{A} mátrixot normálisnak nevezünk, ha $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H$, ahol \mathbf{A}^H jelöli az \mathbf{A} mátrix konjugáltjának transzponáltját.

1.1.18. Tétel. Minden normális mátrix diagonalizálható.

Bizonyítás. Legyen \mathbf{A} egy tetszőleges normális mátrix, λ_1 és \mathbf{v}_1 pedig a mátrix egy sajátpárja. Legyen \mathbf{v}_1 normált, azaz olyan, hogy $\mathbf{v}_1^H\mathbf{v}_1 = 1$. Egészítsük ki ezt a vektort a Gram–Schmidt ortogonalizáció segítségével ortonormált rendszerré (unitér mátrixszá) a $\mathbf{v}_2, \dots, \mathbf{v}_n$ vektorokkal. Ekkor

$$\mathbf{A} \underbrace{[\mathbf{v}_1 \dots \mathbf{v}_n]}_{:=\mathbf{S}_1 \text{ (unitér)}} = [\mathbf{v}_1 \dots \mathbf{v}_n] \begin{bmatrix} \lambda_1 & * & * & \dots \\ 0 & * & * & \dots \\ \vdots & \vdots & \vdots & \\ 0 & * & * & \dots \end{bmatrix},$$

ahonnan, bevezetve az $\mathbf{S}_1 = [\mathbf{v}_1 \dots \mathbf{v}_n]$ és az \mathbf{A}_2 jelölést a jobb oldali második tényező $(2 : n, 2 : n)$ blokkjára, kapjuk, hogy

$$\mathbf{S}_1^H\mathbf{A}\mathbf{S}_1 = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}.$$

Hajtsuk végre az előző eljárást az \mathbf{A}_2 mátrixszal! Ehhez létezik olyan $\tilde{\mathbf{S}}_2$ unitér mátrix, mellyel

$$\tilde{\mathbf{S}}_2^H \mathbf{A}_2 \tilde{\mathbf{S}}_2 = \begin{bmatrix} \lambda_2 & * & * & \dots \\ 0 & * & * & \dots \\ & & \ddots & \\ 0 & * & * & \dots \end{bmatrix}.$$

Legyen

$$\mathbf{S}_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{S}}_2 \end{bmatrix}.$$

Ekkor

$$\mathbf{S}_2^H \mathbf{S}_1^H \mathbf{A} \mathbf{S}_1 \mathbf{S}_2 = \begin{bmatrix} \lambda_1 & * & * & \dots \\ 0 & \lambda_2 & * & \dots \\ 0 & 0 & * & \dots \\ & & \vdots & \vdots \\ 0 & 0 & * & \dots \end{bmatrix}.$$

Hasonlóan folytatva juthatunk el az $\mathbf{S}_3, \dots, \mathbf{S}_{n-1}$ unitér mátrixokhoz, melyekkel

$$\mathbf{S}_{n-1}^H \dots \mathbf{S}_2^H \mathbf{S}_1^H \mathbf{A} \mathbf{S}_1 \mathbf{S}_2 \dots \mathbf{S}_{n-1} = \underbrace{\begin{bmatrix} \lambda_1 & * & * & \dots & * \\ 0 & \lambda_2 & * & \dots & * \\ & & \ddots & & \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}}_{=:\mathbf{T} \text{ (felső háromszög)}}.$$

Legyen $\mathbf{S} = \mathbf{S}_1 \dots \mathbf{S}_{n-1}$. Ez unitér mátrix. Vezessük be a \mathbf{T} jelölést a fenti képletben szereplő felső háromszögmátrixra. Erre a mátrixra igaz, hogy

$$\mathbf{T}^H \mathbf{T} = \mathbf{S}^H \mathbf{A}^H \mathbf{S} \mathbf{S}^H \mathbf{A} \mathbf{S} = \mathbf{S}^H \mathbf{A}^H \mathbf{A} \mathbf{S} = \mathbf{S}^H \mathbf{A} \mathbf{A}^H \mathbf{S} = \mathbf{S}^H \mathbf{A} \mathbf{S} \mathbf{S}^H \mathbf{A}^H \mathbf{S} = \mathbf{T} \mathbf{T}^H,$$

így \mathbf{T} normális mátrix. Mivel \mathbf{T} felső háromszögmátrix, csak úgy lehet normális, ha diagonális. Ebből pedig következik, hogy az \mathbf{A} normális mátrix unitér mátrixszal diagonalizálható. \square

1.1.19. Következmény. A tétel bizonyításából következik, hogy minden \mathbf{A} négyzetes mátrix felírható $\mathbf{A} = \mathbf{S} \mathbf{T} \mathbf{S}^H$ alakban, ahol \mathbf{S} unitér mátrix, \mathbf{T} pedig egy felső háromszögmátrix. Ezt az alakot a mátrixok Schur-felbontásának nevezzük. A hasonlóság miatt a \mathbf{T} mátrix főátlójában az \mathbf{A} mátrix sajátértékei szerepelnek.

1.1.20. Következmény. Az előző tétel következménye, hogy egy mátrix akkor és csak akkor diagonalizálható unitér mátrixszal, ha normális. Az előző tétel bizonyításából kapjuk, hogy a normális mátrixok unitér mátrixszal diagonalizálhatók. A fordított irányhoz tegyük fel, hogy \mathbf{A} unitér mátrixszal diagonalizálható, azaz van olyan \mathbf{S} unitér mátrix, mellyel $\mathbf{S}^H \mathbf{A} \mathbf{S} = \mathbf{\Lambda}$, ahol $\mathbf{\Lambda}$ diagonális mátrix. Ekkor viszont $\mathbf{A} = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^H$, és

$$\begin{aligned} \mathbf{A}^H \mathbf{A} &= (\mathbf{S} \mathbf{\Lambda} \mathbf{S}^H)^H (\mathbf{S} \mathbf{\Lambda} \mathbf{S}^H) = \mathbf{S} \mathbf{\Lambda}^H \mathbf{S}^H \mathbf{S} \mathbf{\Lambda} \mathbf{S}^H = \mathbf{S} \mathbf{\Lambda}^H \mathbf{\Lambda} \mathbf{S}^H = \\ &= \mathbf{S} \mathbf{\Lambda} \mathbf{\Lambda}^H \mathbf{S}^H = (\mathbf{S} \mathbf{\Lambda} \mathbf{S}^H) (\mathbf{S} \mathbf{\Lambda}^H \mathbf{S}^H) = \mathbf{A} \mathbf{A}^H. \end{aligned}$$

1.1.21. Tétel. *Egy valós mátrix akkor és csak akkor diagonalizálható ortogonális mátrixszal, ha szimmetrikus.*

Bizonyítás. Először belátjuk, hogy ha egy \mathbf{A} valós mátrix ortogonális mátrixszal diagonalizálható, akkor az szimmetrikus. Legyen \mathbf{S} ortogonális és $\mathbf{A} = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T$. Ekkor $\mathbf{A}^T = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T = \mathbf{A}$, azaz \mathbf{A} szimmetrikus.

A fordított irány igazolásához tudjuk, hogy a szimmetrikus mátrixok normálisak, ezért diagonalizálhatók. Így van lineárisan független sajátvektorrendszerük. Azt kell csak belátnunk, hogy ezek a vektorok választhatók ortonormáltan. Ha egy sajátértékhez több lineárisan független sajátvektor is tartozik, akkor ezek a Gram–Schmidt eljárással ortonormálhatók. Már csak azt kell megmutatnunk, hogy a különböző sajátértékekhez tartozó sajátvektorok nemcsak függetlenek, hanem ortogonálisak is. Legyen \mathbf{v}_λ és \mathbf{v}_μ két különböző sajátértékhez (λ és μ) tartozó sajátvektor. Mivel a mátrix szimmetrikus, így a sajátértékei és a sajátvektorai is valósak. A szimmetria miatt a

$$\begin{aligned} \mathbf{v}_\lambda^T \mathbf{A} \mathbf{v}_\mu &= \mathbf{v}_\lambda^T \mu \mathbf{v}_\mu = \mu \mathbf{v}_\lambda^T \mathbf{v}_\mu, \\ \mathbf{v}_\mu^T \mathbf{A} \mathbf{v}_\lambda &= \mathbf{v}_\mu^T \lambda \mathbf{v}_\lambda = \lambda \mathbf{v}_\mu^T \mathbf{v}_\lambda = \lambda \mathbf{v}_\lambda^T \mathbf{v}_\mu \end{aligned}$$

értékeknek meg kell egyezniük. Ez pedig csak úgy lehetséges, ha $\mathbf{v}_\lambda^T \mathbf{v}_\mu = 0$, azaz a különböző sajátértékekhez tartozó sajátvektorok ortogonálisak. Így választható ortonormált sajátvektorrendszer. A mátrix azzal az ortogonális mátrixszal diagonalizálható, melynek oszlopai az ortonormált sajátvektorok. \square

1.1.3. Sajátérték-feladatok kondicionáltsága

Korábban láttuk, hogy a sajátérték-feladat megoldható a karakterisztikus egyenlet segítségével. Ezzel csak az a baj, hogy négyenél nagyobb fokszámú polinomok megoldására

nincs megoldóképletünk, így sokszor nem jutunk eredményre ezzel a módszerrel. Éppen ezért a sajátértékeket általában direkt módszerrel nem lehet meghatározni. Ezért a gyakorlatban szinte mindig iterációs módszereket alkalmazunk. Ekkor elegendő a megfelelő sajátértéket vagy sajátvektort közelíteni, mert az egyik közelítésének ismeretében a másik közelítését könnyen meg lehet oldani.

Mielőtt rátérnénk a sajátérték-feladatok numerikus megoldásának lehetőségeire, vizsgáljuk meg a sajátérték-feladat kondicionáltságát! Mennyit változnak az \mathbf{A} négyzetes mátrix sajátértékei, ha a mátrix elemeit kicsit megváltoztatjuk?

1.1.22. Tétel. (Bauer–Fike, 1960) *Legyen $\mathbf{A} \in \mathbb{R}^{n \times n}$ egy diagonalizálható mátrix, azaz \mathbf{A} felírható $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ alakban, ahol az egyszerűség kedvéért feltesszük, hogy \mathbf{V} oszlopai normálva vannak a $\|\cdot\|_p$ vektornormában, és \mathbf{D} diagonális mátrix. Legyen $\delta\mathbf{A}$ egy tetszőleges mátrix, és legyen μ az $\mathbf{A} + \delta\mathbf{A}$ mátrix egy sajátértéke. Ekkor*

$$\min_{\lambda \text{ \mathbf{A} sajátértéke}} |\lambda - \mu| \leq \kappa_p(\mathbf{V}) \cdot \|\delta\mathbf{A}\|_p,$$

ahol κ_p a kondíciószámot jelöli.

Bizonyítás. Ha μ sajátértéke \mathbf{A} -nak is, akkor triviális az állítás. Tegyük fel, hogy nem sajátértéke. Mivel $\mathbf{A} + \delta\mathbf{A} - \mu\mathbf{I}$ szinguláris, ezért

$$\mathbf{V}^{-1}(\mathbf{A} + \delta\mathbf{A} - \mu\mathbf{I})\mathbf{V} = \mathbf{D}\mathbf{V}^{-1}\delta\mathbf{A}\mathbf{V} - \mu\mathbf{I}$$

is szinguláris, így van olyan $\mathbf{x} \neq \mathbf{0}$ vektor, mellyel

$$(\mathbf{D} - \mu\mathbf{I} + \mathbf{V}^{-1}\delta\mathbf{A}\mathbf{V})\mathbf{x} = \mathbf{0}.$$

A $\mathbf{D} - \mu\mathbf{I}$ mátrix inverzével balról szorozva azt kapjuk, hogy

$$(\mathbf{I} + (\mathbf{D} - \mu\mathbf{I})^{-1}\mathbf{V}^{-1}\delta\mathbf{A}\mathbf{V})\mathbf{x} = \mathbf{0},$$

azaz

$$\mathbf{x} = -(\mathbf{D} - \mu\mathbf{I})^{-1}\mathbf{V}^{-1}\delta\mathbf{A}\mathbf{V}\mathbf{x}.$$

Így

$$\|\mathbf{x}\|_p \leq \|(\mathbf{D} - \mu\mathbf{I})^{-1}\|_p \cdot \|\mathbf{V}^{-1}\|_p \cdot \|\delta\mathbf{A}\|_p \cdot \|\mathbf{V}\|_p \cdot \|\mathbf{x}\|_p$$

és

$$1 \leq \|(\mathbf{D} - \mu\mathbf{I})^{-1}\|_p \cdot \kappa_p(\mathbf{V}) \cdot \|\delta\mathbf{A}\|_p = \max_i \frac{1}{|\lambda_i - \mu|} \kappa_p(\mathbf{V}) \cdot \|\delta\mathbf{A}\|_p = \frac{1}{\min_i |\lambda_i - \mu|} \kappa_p(\mathbf{V}) \cdot \|\delta\mathbf{A}\|_p.$$

Ebből pedig következik az állítás. \square

1.1.23. Megjegyzés. A sajátérték-feladat kondicionáltságát nem az eredeti mátrix, hanem a diagonalizáló mátrix (a sajátvektorok mátrixa) kondíciószáma határozza meg. Ez speciálisan azt jelenti, hogy mivel a szimmetrikus mátrixok ortogonális mátrixokkal diagonalizálhatók, és az ortogonális mátrixok 2-es normája 1, ezért szimmetrikus mátrixokra a

$$\min_{\lambda \in \text{A sajátértéke}} |\lambda - \mu| \leq \|\delta \mathbf{A}\|_2$$

becslés érvényes.

1.2. Motiváció

1.2.1. Stabilitási problémák

Legyen adott egy mindkét végén befogott rúd, melynek felső végét P erő terheli. Ekkor a nyugalmi helyzetből való $u(x)$ kismértékű kitérés közelítőleg eleget tesz a

$$JEu'' = -Pu, \quad 0 < x < 1 \quad (1.1)$$

differenciálegyenletnek. Itt u'' az x hosszúsági koordináta szerinti második deriváltat jelöli, JE pedig megadja a rúd rugalmassági tulajdonságait (E a rúd anyagát jellemző Young-féle modulus, J a rúd keresztmetszetétől függő axiális nyomaték.)

A csuklós befogás miatt kellenek az

$$u(0) = 0, \quad u(1) = 0 \quad (1.2)$$

feltételeket. Az (1.1), (1.2) egyenleteknek mindig van megoldása, hiszen az $u(x) \equiv 0$ megoldás, ami azt jelenti, hogy a nyugalmi helyzetben van a rúd. Ez akkor változik meg, ha (1.1), (1.2)-nek vannak nullától különböző megoldásai. (1.1) miatt kell, hogy ezek a megoldások

$$u = a \sin \omega x + b \cos \omega x, \quad \text{ahol } \omega := \lambda^{\frac{1}{2}}, \quad \lambda := \frac{P}{JE} \quad (1.3)$$

alakúak legyenek és (1.2)-ből pedig kapjuk, hogy $b = 0$ és $\omega = k\pi$, $k = 1, 2, \dots$, azaz a keresett nemtriviális megoldások csak a

$$\lambda = \lambda_k = (k\pi)^2, \quad k = 1, 2, \dots, \quad (1.4)$$

értékek, azaz a

$$P_k = (k\pi)^2 JE, \quad k = 1, 2, \dots$$

erők mellett jöhetnek létre. A legkisebb ilyen erő az ún. törőerő:

$$P_1 = P_t = \pi^2 JE.$$

Itt most a (1.1) differenciálegyenletre vonatkozó sajátérték-feladatot oldottuk meg a (1.2) feltételek mellett. Ekkor áttérhetünk szokásos sajátérték-feladatra, elég a differenciálegyenletet lineáris rendszerrel helyettesíteni. Ekkor a \mathbf{v} vektorra – amelynek \mathbf{v}_i komponense az $u(\frac{i}{n+1})$ érték közelítése, $i = 1, 2, \dots$ – az

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0} \quad (1.5)$$

standard alakú sajátérték-feladatot (ahol $\lambda = \frac{P}{JE}$) kapjuk. Itt az \mathbf{A} mátrix legyen a következő:

$$\mathbf{A} := \frac{1}{h^2} \begin{bmatrix} 2 & -1 & 0 & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot \\ 0 & -1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & -1 & 2 & -1 \\ 0 & \cdot & \cdot & \cdot & -1 & 2 \end{bmatrix} = (n+1)^2 \text{tridiag}(-1, 2, -1).$$

Az \mathbf{A} mátrix szimmetrikus, ami általában a mechanikai eredetű feladatokra jellemző. Az \mathbf{A} mátrix speciális alakját és a

$$\begin{aligned} \sin k\pi(x_i + h) + \sin k\pi(x_i - h) &= 2 \sin k\pi x_i \cos k\pi h, \\ 2 - 2 \cos k\pi h &= 4 \sin^2 \frac{k\pi h}{2} \end{aligned}$$

azonosságokat (ahol $x_i := ih$, $h := \frac{1}{n+1}$) felhasználva azt kapjuk, hogy

$$\begin{aligned} \mathbf{A}\mathbf{v}^k &= \lambda_k^h \mathbf{v}^k, \quad k = 1, 2, \dots, n, \\ \lambda_k^h &:= \frac{4}{h^2} \sin^2 \frac{k\pi h}{2}, \quad \mathbf{v}_i^{(k)} := \sin k\pi x_i, \quad i = 1, \dots, n. \end{aligned} \quad (1.6)$$

Ez az ún. teljes sajátérték-feladat megoldása: megtaláltuk \mathbf{v}^k sajátvektorokat (az (1.5) rendszer nemtriviális megoldásait) és a λ_k^h sajátértékeket.

A $\{\mathbf{v}^k\}$ sajátvektorok ortogonális rendszert alkotnak, mivel ha egy szimmetrikus mátrixnak minden sajátértéke különböző, akkor sajátvektorai ortogonális rendszert alkotnak, azaz

$$(\mathbf{v}^k, \mathbf{v}^l) := \sum_{j=1}^n \sin \frac{k\pi j}{n+1} \sin \frac{l\pi j}{n+1} = 0, \quad k \neq l,$$

1.2.2. Internetes oldalak rangsorolása

Az internetes oldalakat rangsoroló algoritmusnak fontos része az oldalak és linkjeik által definiált nemnegatív mátrix abszolút értékben maximális sajátértékéhez tartozó sajátvektor meghatározása, mivel annak (nemnegatív) komponensei adják a rangsort. Jelölje \mathbf{A} ezt a mátrixot.

Legyen n az internetoldalak aktuális száma (ez több milliárd is lehet). Továbbá legyen $\mathbf{M} = (m_{ij}) \in \mathbb{R}^{n \times n}$ az incidenciamátrix:

$$m_{ij} = \begin{cases} 1, & \text{ha az } i\text{-edik oldalra mutat link a } j\text{-edikből} \\ 0, & \text{különben.} \end{cases}$$

Legyen $N(j)$ a j -edik oldalból induló linkek száma. Ekkor $\mathbf{A} = \text{diag}_j(N(j))^+$, ahol a diagonális mátrix általánosított inverzét a „+” jelöli.

Ez a mátrix nemnegatív, ezért van nemnegatív sajátértéke, és a hozzá tartozó sajátvektor is nemnegatív, ezt tudjuk Frobenius tételéből. Ezen sajátvektor maximális komponensének indexe adja az első internetoldalt, a második legnagyobb komponens indexe a második oldalt stb. A sajátvektort a rangsoroló algoritmus a hatványmódszerrel határozza meg (ld. későbbi fejezetekben) normálás nélkül.

A fenti \mathbf{A} mátrixszal az a baj, hogy nem mindig vezet elfogadható eredményekre (vagy a mátrix maximális sajátértéke nem egyszeres, így a hatványmódszer nem konvergens, vagy az adódó vektor nem alkalmas a fontos oldalak meghatározására), ezért \mathbf{A} helyett vizsgáljuk a következő mátrixot:

$$\mathbf{A}_\varepsilon := (1 - \varepsilon)\mathbf{A} + \frac{\varepsilon}{n}\mathbf{E}, \quad \text{ahol } \mathbf{E} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Ennek a mátrixnak (mint irreducibilis pozitív mátrixnak) már Frobenius tétele szerint az abszolút értékben maximális sajátértéke (pozitív és) egyszeres, a hozzátartozó sajátvektor minden komponense pozitív, és ehhez konvergál a hatványmódszer.

2. fejezet

A sajátértékeket egyenként közelítő eljárások

A numerikus matematikában a sajátérték-meghatározási módszereket két csoportra szokás osztani. A sajátértékeket egyenként közelítő eljárásokra, illetve a sajátértékeket egyszerre közelítőkre. Az első csoportba tartozó módszerek csak egy-egy megfelelő sajátértéket közelítenek, míg a második csoportba tartozó eljárások egyszerre az összes sajátértékre adnak közelítést. Kezdjük a sajátértékeket egyenként közelítő eljárásokkal.

A fejezetben a hatványmódszer és az inverz iteráció rész [3], a Householder-eljárás [1], és a hatványmódszerig tartó rész pedig [4] alapján készült el.

A sajátértékeket egyenként közelítő eljárások egy olyan vektorsorozatot állítanak elő, amely egy meghatározott sajátvektorhoz tart. A sajátvektort ekkor ezen sorozat egy határértékhez elegendően közeli elemével közelítjük.

2.0.1. Tétel. *Legyen adott egy $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ vektor és az $\mathbf{A} \in \mathbb{R}^{n \times n}$ mátrix. Ekkor*

$$\min_{a \in \mathbb{R}} \|\mathbf{Ax} - a\mathbf{x}\|_2^2 = \|\mathbf{Ax} - \rho(\mathbf{x})\mathbf{x}\|_2^2,$$

ahol

$$\rho(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}}.$$

Bizonyítás. A bal oldalon egy a -tól függő egyváltozós függvény áll. Számoljuk ki ennek az értékét!

$$\begin{aligned} \|\mathbf{Ax} - a\mathbf{x}\|_2^2 &= (\mathbf{x}^T \mathbf{A}^T - a\mathbf{x}^T)(\mathbf{Ax} - a\mathbf{x}) = \mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - 2a\mathbf{x}^T \mathbf{Ax} + a^2 \mathbf{x}^T \mathbf{x} = \\ &= a^2 \mathbf{x}^T \mathbf{x} - 2a\mathbf{x}^T \mathbf{Ax} + \mathbf{x}^T \mathbf{A}^T \mathbf{Ax}. \end{aligned}$$

Mivel $\mathbf{x}^T \mathbf{x} > 0$, ha $\mathbf{x} \neq 0$, ezért a függvény a minimumát az

$$a_{\min} = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \rho(\mathbf{x})$$

esetben veszi fel. \square

2.0.2. Definíció. Legyen $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ vektor és az $\mathbf{A} \in \mathbb{R}^{n \times n}$ mátrix. Ekkor az

$$\rho(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

számot az \mathbf{x} vektorhoz tartozó Rayleigh-hányadosnak hívjuk.

A tétel szerint tehát 2-es normában egy vektor Rayleigh-hányadosszorosa lesz legközelebb \mathbf{x} számszorosai közül az $\mathbf{A} \mathbf{x}$ vektorhoz.

2.0.3. Megjegyzés. Szimmetrikus mátrixok esetén a Rayleigh-hányados mindig a legkisebb és a legnagyobb sajátérték között helyezkedik el, azaz

$$\lambda_{\min} \leq \rho(\mathbf{x}) \leq \lambda_{\max},$$

továbbá

$$\lambda_{\max} = \max_{\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n} \rho(\mathbf{x}), \quad \lambda_{\min} = \min_{\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n} \rho(\mathbf{x}).$$

Ha tehát van egy közelítésünk a sajátvektorra, akkor a Rayleigh-hányados segítségével kaphatunk jó becslést a sajátvektorhoz tartozó sajátértékre.

2.1. A hatványmódszer

Ez a módszer az abszolút értékben legnagyobb sajátértéket és a hozzá tartozó sajátvektort állítja elő. Minden iterációs lépésben kiszámít egy új vektort, és ezen sorozat vektoraival közelíti a keresett sajátvektort, a hozzá tartozó sajátérték az iteráció melékterméke.

Jelöljük a mátrix sajátértékeit λ_k -val és sajátvektorait $\mathbf{v}^{(k)}$ -val.

Nézzük meg először a módszer működését egy kisebb példán. Indulunk ki valamilyen $\mathbf{y}^0 \neq \mathbf{0}$ vektorból és képzeljük el először, hogy $\lambda_1 = \frac{1}{10}$, $\lambda_2 = 1$, $\lambda_3 = 10$, valamint, hogy

$$\mathbf{y}^0 = a_1 \mathbf{v}^{(1)} + a_2 \mathbf{v}^{(2)} + a_3 \mathbf{v}^{(3)},$$

ahol $a_i \neq 0$, $i = 1, 2, 3$.

Most szorozzuk \mathbf{y}^0 -t \mathbf{A} -val, felhasználva, hogy $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, ahol $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{v} \neq \mathbf{0}$ vektor és $\lambda \in \mathbb{R}$:

$$\begin{aligned}\mathbf{y}^1 &:= \mathbf{A}\mathbf{y}^0 = \mathbf{A}(a_1\mathbf{v}^{(1)} + a_2\mathbf{v}^{(2)} + a_3\mathbf{v}^{(3)}) = a_1\mathbf{A}\mathbf{v}^{(1)} + a_2\mathbf{A}\mathbf{v}^{(2)} + a_3\mathbf{A}\mathbf{v}^{(3)} = \\ &= a_1\lambda_1\mathbf{v}^{(1)} + a_2\lambda_2\mathbf{v}^{(2)} + a_3\lambda_3\mathbf{v}^{(3)} = \frac{1}{10}a_1\mathbf{v}^{(1)} + a_2\mathbf{v}^{(2)} + 10a_3\mathbf{v}^{(3)},\end{aligned}$$

Ezután ismételjük meg ezt a műveletet \mathbf{y}^1 -gyel:

$$\begin{aligned}\mathbf{y}^2 &:= \mathbf{A}\mathbf{y}^1 = \mathbf{A}\left(\frac{1}{10}a_1\mathbf{v}^{(1)} + a_2\mathbf{v}^{(2)} + 10a_3\mathbf{v}^{(3)}\right) = \\ &= \frac{1}{10}a_1\mathbf{A}\mathbf{v}^{(1)} + a_2\mathbf{A}\mathbf{v}^{(2)} + 10a_3\mathbf{A}\mathbf{v}^{(3)} = \frac{1}{10^2}a_1\mathbf{v}^{(1)} + a_2\mathbf{v}^{(2)} + 10^2a_3\mathbf{v}^{(3)}.\end{aligned}$$

Ismételjük összesen m -szer:

$$\mathbf{y}^m := \mathbf{A}\mathbf{y}^{m-1} = \frac{1}{10^m}a_1\mathbf{v}^{(1)} + a_2\mathbf{v}^{(2)} + 10^m a_3\mathbf{v}^{(3)}. \quad (2.1)$$

Ekkor jól látható, hogy ebben az összegben az első tagot elhanyagolhatjuk, sőt a másodikat is, legalábbis az utolsóhoz képest – mégpedig akkor is, ha az a_i -k nagyságrendje eléggé különböző. Az \mathbf{y}^m vektorok növekvő m -mel egyre közelebb kerülnek $\mathbf{v}^{(3)}$ egy többszöröséhez, és csak a túlsordulást kellene megakadályozni. Ezt úgy lehet megakadályozni, hogy néha, vagy akár minden lépésben \mathbf{y}^m -ről $\frac{\mathbf{y}^m}{\|\mathbf{y}^m\|}$ -ra megyünk át valamilyen vektornormát alkalmazva:

$$\mathbf{y}^0 \neq \mathbf{0}, m = 1, 2, \dots : \mathbf{x} := \mathbf{A}\mathbf{y}^{m-1}, \mathbf{y}^m = \frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

2.1.1. Konvergenciafeltételek

Három olyan feltétel van, amelyek biztosítják, hogy az előbb említett helyzetbe kerüljünk, amikor az $\{\mathbf{y}^m\}_{m=0}^\infty$ sorozat konvergál a legnagyobb sajátértékhez tartozó sajátvektorhoz. Ezek a következők:

1. Az \mathbf{A} mátrix legyen diagonalizálható, pl. normális.
2. Legyen $(\mathbf{y}^0, \mathbf{v}^{(0)}) \neq 0$, ahol $\mathbf{v}^{(0)}$ a keresett sajátvektor.
3. Teljesüljön, hogy

$$|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_{n-1}| < |\lambda_n|. \quad (2.2)$$

Ha az első feltétel teljesül, akkor tudjuk, hogy tetszőleges \mathbf{y}^0 vektort mindig ki tudunk fejezni a sajátvektorok lineáris kombinációjaként.

A második feltétel azt biztosítja, hogy \mathbf{y}^0 tartalmazza a keresett sajátvektort. Természetesen itt az euklideszi skalárszorzat segítségével van megfogalmazva a feltétel.

Végül a harmadik feltétel garantálja, hogy λ_n hatványai növekvő hatványkitevővel abszolút értékben egyre jobban különbözni fognak a többi sajátérték hatványaitól.

Ellenőrizzük, hogy a három feltétel valóban elegendő a konvergenciához abban az esetben, amikor \mathbf{A} normális, vagyis sajátvektorai ortonormált rendszert alkotnak. Az euklideszi skalárszorzatban érvényes

$$(\mathbf{v}^{(i)}, \mathbf{v}^{(j)}) = \delta_{ij}, \quad 1 \leq i, \quad j \leq n.$$

Első feltételből tudjuk, hogy \mathbf{y}^0 előáll a következő alakban:

$$\mathbf{y}^0 = \sum_{i=1}^n c_i \mathbf{v}^{(i)}. \quad (2.3)$$

Határozzuk meg a lineáris kombináció c_i együtthatóit. Ehhez az (2.3) relációt skalárisan szorozzuk meg $\mathbf{v}^{(j)}$ -vel valamelyik j -re, $1 \leq j \leq n$:

$$(\mathbf{y}^0, \mathbf{v}^{(j)}) = \left(\sum_{i=1}^n c_i \mathbf{v}^{(i)}, \mathbf{v}^{(j)} \right) = \sum_{i=1}^n c_i (\mathbf{v}^{(i)}, \mathbf{v}^{(j)}) = \sum_{i=1}^n c_i \delta_{ij} = c_j. \quad (2.4)$$

Számítsuk ki az \mathbf{y}^0 euklideszi normáját:

$$\|\mathbf{y}^0\|_2^2 = (\mathbf{y}^0, \mathbf{y}^0) = \left(\sum_{i=1}^n c_i \mathbf{v}^{(i)}, \mathbf{y}^0 \right) = \sum_{i=1}^n c_i (\mathbf{v}^{(i)}, \mathbf{y}^0).$$

Az \mathbf{y}^0 (2.3) lineáris kombinációjában az összegzési indexet változtassuk j -re, mivel az i már szerepelt, és ezzel le van kötve. Így

$$\|\mathbf{y}^0\|_2^2 = \sum_{i=1}^n c_i (\mathbf{v}^{(i)}, \mathbf{y}^0) = \sum_{i=1}^n c_i (\mathbf{v}^{(i)}, \sum_{j=1}^n c_j \mathbf{v}^{(j)}) = \sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j (\mathbf{v}^{(i)}, \mathbf{v}^{(j)}).$$

Innen végül azt kapjuk, hogy

$$\|\mathbf{y}^0\|_2^2 = \sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j (\mathbf{v}^{(i)}, \mathbf{v}^{(j)}) = \sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j \delta_{ij} = \sum_{i=1}^n |c_i|^2. \quad (2.5)$$

Ezután $\mathbf{A}\mathbf{y}^0$ -ra kapjuk, hogy

$$\mathbf{A}\mathbf{y}^0 = \sum_{i=1}^n c_i \mathbf{A}\mathbf{v}^{(i)} = \sum_{i=1}^n c_i \lambda_i \mathbf{v}^{(i)}.$$

Ha most (2.5)-ban \mathbf{y}^0 helyett $\mathbf{A}\mathbf{y}^0$ írunk és c_i helyett $c_i\lambda_i$ -t, akkor az eredmény

$$\|\mathbf{A}\mathbf{y}^0\|_2^2 = \sum_{i=1}^n |c_i\lambda_i|^2.$$

Hasonlóan számítjuk ki, hogy

$$\mathbf{A}^m\mathbf{y}^0 = \sum_{i=1}^n c_i\lambda_i^m\mathbf{v}^{(i)} \text{ és } \|\mathbf{A}^m\mathbf{y}^0\|_2^2 = \sum_{i=1}^n |c_i\lambda_i^m|^2,$$

tehát

$$\|\mathbf{A}^m\mathbf{y}^0\|_2 = \left[\sum_{i=1}^n |c_i\lambda_i^m|^2 \right]^{\frac{1}{2}}.$$

Idáig csak az első konvergenciafeltételt használtuk. A második feltételből és (2.4)-ből tudjuk, hogy $c_n \neq 0$. Emeljük ki az összegből az n -edik tagot, hogy a harmadik feltételt is tudjuk használni:

$$\mathbf{A}^m\mathbf{y}^0 = \lambda_n^m \left[c_n\mathbf{v}^{(n)} + \sum_{i=1}^{n-1} c_i \left(\frac{\lambda_i}{\lambda_n} \right)^m \mathbf{v}^{(i)} \right],$$

$$\|\mathbf{A}^m\mathbf{y}^0\|_2 = |\lambda_n|^m \left[|c_n|^2 + \sum_{i=1}^{n-1} |c_i|^2 \left| \frac{\lambda_i}{\lambda_n} \right|^{2m} \right]^{\frac{1}{2}}.$$

Ekkor nagy m estén azt kapjuk, hogy

$$\mathbf{A}^m\mathbf{y}^0 = \lambda_n^m \left[c_n\mathbf{v}^{(n)} + O\left(\frac{\lambda_{n-1}}{\lambda_n}\right)^m \right] \approx \lambda_n^m c_n \mathbf{v}^{(n)},$$

$$\|\mathbf{A}^m\mathbf{y}^0\|_2 = |\lambda_n|^m \left[|c_n|^2 + O\left(\left|\frac{\lambda_{n-1}}{\lambda_n}\right|^{2m}\right) \right]^{\frac{1}{2}} \approx |\lambda_n|^m |c_n|.$$

Tehát

$$\frac{\mathbf{A}^m\mathbf{y}^0}{\|\mathbf{A}^m\mathbf{y}^0\|_2} \approx \kappa_m \mathbf{v}^{(n)},$$

ahol $\kappa_m = \frac{\lambda_n^m c_n}{|\lambda_n^m c_n|}$ olyan komplex szám, amelynek abszolút értéke 1. Ettől a κ_m skalárszorozattól eltekintve pontosan megkaptuk a keresett sajátvektort, de sajátvektoroknál egy nemnulla skalárszorozó nem lényeges: eredménynek $\kappa_m \mathbf{v}^{(n)}$ ugyanolyan jó, mint $\mathbf{v}^{(n)}$.

A konvergencia sebessége $\left(\frac{\lambda_{n-1}}{\lambda_n}\right)$ -vel jellemezhető. Ha a $\left|\frac{\lambda_{n-1}}{\lambda_n}\right|$ hányados közel van 1-hez, akkor a konvergencia lassú.

Most lássuk a három konvergenciafeltétel fontosságát.

Az iteráció akkor is konvergál, ha \mathbf{A} nem normális vagy diagonalizálható, ezért az első nem annyira fontos, de a számításokat megkönnyítette. A lényeges feltétel a harmadik, amelyet másképpen úgy szokás mondani, hogy λ_n -nek dominánsnak kell lennie.

Lássunk egy példát arra, hogy ha az első két feltétel teljesül, de a harmadik nem, akkor a módszer valóban nem lesz konvergens. Legyen

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \text{ és } \mathbf{y}^0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}. \quad (2.6)$$

Ez a mátrix szimmetrikus, tehát diagonalizálható, és $\lambda_{1,2} = \pm 1$, az ortonormált sajátvektorok

$$\mathbf{v}^{(1)} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}^{(2)} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

A megadott \mathbf{y}^0 esetén a második feltétel is teljesül: $(\mathbf{y}^0, \mathbf{v}^{(2)}) = -\frac{\sqrt{2}}{2}$. De a harmadik feltétel nem teljesül, hiszen $|\lambda_1| = |\lambda_2|$. Az $\{\mathbf{y}^m\}$ sorozat nem konvergál: páros m -re $\mathbf{y}^m = \mathbf{y}^0$, páratlan m -re viszont $\mathbf{y}^m = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, mert \mathbf{A} csak permutálja a vele megszorozott vektorok komponenseit, és a szorzatban végtelen sokszor szereplő két vektor sem sajátvektor.

Gyakran azért nem teljesül $|\lambda_{n-1}| < |\lambda_n|$, mert \mathbf{A} valós és $\lambda_{n-1} = \overline{\lambda_n}$.

Most foglalkozzunk a második feltétellel. Ha $(\mathbf{y}^0, \mathbf{v}^{(n)}) = 0$ vagy kicsi, akkor az iteráció folyamán (kerekítési hibák hatására) keletkezik egy $\mathbf{v}^{(n)}$ -irányú komponens az \mathbf{y}^m vektorokban. Ekkor késleltetve ugyan, de végül mégis $\mathbf{v}^{(n)}$ -hez konvergál az iteráció. Hogy ez mikor érvényesül az az alkalmazott pontosságtól függ, továbbá attól, hogy λ_n mennyire dominál, végül a többi c_i együttható nagysága is befolyásolja.

Mivel $\mathbf{v}^{(n)}$ -t és λ_n -t keressük, így ezek nyilván ismeretlenek, azonban az első és a második feltételek ezekre vonatkoznak. Tehát próbálkozni kell. Ha nem kapunk konvergenciát egy adott kezdeti vektorral, akkor egy másikat kell választanunk és azzal tovább próbálkozni.

A hatványmódszerrel nemcsak a sajátvektort, hanem a hozzá tartozó sajátértéket is megkaphatjuk a következő módszerekkel:

1. Ha $\mathbf{y}_i^{(m)} \neq 0$, akkor $\frac{(\mathbf{A}\mathbf{y}^m)_i}{\mathbf{y}_i^{(m)}} \approx \lambda_n$ nagy m -re.
2. $\|\mathbf{A}\mathbf{y}^m\| \rightarrow |\lambda_n|$, $m \rightarrow \infty$; ezt kombináljuk az előbbi gondolattal.

3. A legcélszerűbb az, ha a Rayleigh-hányadost használjuk: $\|\mathbf{Ax} - \rho\mathbf{x}\|$ akkor lesz minimális (itt továbbra is euklideszi norma van), ha

$$\rho(\mathbf{x}) = \frac{(\mathbf{Ax}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}.$$

Amennyiben most \mathbf{x} az \mathbf{A} sajátvektora, akkor $\rho(\mathbf{x})$ lesz a hozzá tartozó sajátérték.

2.1.2. A hatványmódszer algoritmus

Adott n , az $\mathbf{A} \in \mathbb{R}^{n \times n}$ mátrix, egy $\mathbf{x} \neq \mathbf{0}$ kezdeti vektor, a *maxit* maximális iterációs szám, valamint egy ε pontosság. Az algoritmusban kiszámításra kerülő μ_m számok reményeink szerint növekvő m -mel a λ_n maximális abszolút értékű sajátértékhez közelednek, és ilyen siker esetén \mathbf{x} lesz a hozzá tartozó közelítő sajátvektor.

0. $m := 0$, $\mathbf{x}_n := \|\mathbf{x}\|_2$, ? $\mathbf{x}_n = \mathbf{0}$? [stop: „Adj nemnulla kezdeti vektort!”]
1. $\mathbf{x} := \frac{\mathbf{x}}{\mathbf{x}_n}$, $\mathbf{y} := \mathbf{A} * \mathbf{x}$, $\mu_m := (\mathbf{y}, \mathbf{x})$
2. $m := 1(1)maxit$ % m fut 1-től egyesével *maxit*-ig
3. [$\mathbf{y}_n := \|\mathbf{y}\|_2$, ? $\mathbf{y}_n = \mathbf{0}$? [stop: \mathbf{x} „sajátvektor”, $\lambda = 0$ „sajátérték”]
4. $\mathbf{x} := \frac{\mathbf{y}}{\mathbf{y}_n}$ a normálás miatt $\|\mathbf{x}\| = 1$ lesz
5. $\mathbf{y} = \mathbf{A} * \mathbf{x}$, $\mu_m := (\mathbf{y}, \mathbf{x})$
6. ? $|\mu_m - \mu_{m-1}| \leq \varepsilon(1 + |\mu_m|)$? [goto 9.]]_m
7. [stop: „Elértük a maximális iterációs számot, jelenlegi eredmény:” \mathbf{x} „vektor”, μ_m „érték”.]
8. ? $\|\mathbf{y} - \mu_m * \mathbf{x}\|_2^2 \leq \varepsilon$? [stop: „Megtaláltam a keresett adatokat!” \mathbf{x} „közelítő sajátvektor”, μ_m „közelítő sajátérték”, m „iteráció”]
9. [stop: „Nem találtam megoldást! Jelenlegi eredmény:” \mathbf{x} „vektor”, μ_m „érték”]

Fontos, hogy a 8. lépés tesztje a norma négyzetére vonatkozik.

2.1.3. Az eltolás

Eddig a hatványmódszerrel csak az abszolút értékben legnagyobb sajátértéket és a hozzá tartozó sajátvektort kaptuk meg.

Ezen tudunk segíteni eltolással. Ez abból áll, hogy \mathbf{A} helyett az $\mathbf{A} - \lambda_0 \mathbf{I}$ mátrixot vizsgáljuk, ahol a λ_0 számot alkalmasan fogjuk megadni. Ebben az esetben ugyanis, az $\mathbf{A} - \lambda_0 \mathbf{I}$ mátrix sajátvektorai megegyeznek az \mathbf{A} sajátvektoraival, de a sajátértékek megváltoznak, a következő módon: Ha

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{2.7}$$

teljesül, ahol $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{v} \neq \mathbf{0}$ vektor és $\lambda \in \mathbb{R}$, akkor

$$(\mathbf{A} - \lambda_0 \mathbf{I})\mathbf{x} = \lambda\mathbf{x} - \lambda_0\mathbf{x} = (\lambda - \lambda_0)\mathbf{x}.$$

Innen látjuk, hogy éppen λ_0 -val csúsztak el a sajátértékek. Ha most az $\mathbf{A} - \lambda_0 \mathbf{I}$ mátrixra alkalmazzuk a hatványmódszert, akkor a sajátértékek λ_i helyett $\lambda_i - \lambda_0$ lesznek, és ekkor már nem feltétlenül a korábban domináns λ_n lesz most is domináns.

Legyen \mathbf{A} például szimmetrikus és pozitív definit, $n > 1$. Ekkor a sajátértékei a $[\lambda_1, \lambda_n]$ intervallumban vannak, és a mátrix spektrálsugara $\lambda_n = \rho(\mathbf{A})$, valamint

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n,$$

és tegyük fel, hogy $\lambda_1 < \lambda_n$. Legyen $\lambda_0 = \|\mathbf{A}\|_1$ -nek, mivel ez könnyen kiszámítható. Ekkor $\mathbf{A} - \lambda_0 \mathbf{I}$ sajátértékei a $[\lambda_1 - \lambda_0, \lambda_n - \lambda_0]$ intervallumban vannak:

$$\lambda_1 - \lambda_0 \leq \lambda_2 - \lambda_0 \leq \dots \leq \lambda_n - \lambda_0 \leq 0,$$

mert $\lambda_n = \rho(\mathbf{A}) \leq \|\mathbf{A}\|_1 = \lambda_0$. Vagyis az abszolút értékben legkisebb sajátérték az eltolás révén az abszolút értékben legnagyobb sajátérték lesz.

A kiszámított közelítő sajátértéket μ_1 -gyel jelölve, érvényes $\mu_1 \approx \lambda_1 - \lambda_0$, és innen a $\lambda_1 \approx \mu_1 + \lambda_0$ értéket kapjuk. Ezzel az az egyetlen baj, hogy ha a legnagyobb és legkisebb sajátérték néhány nagyságrendben eltér egymástól, akkor a $\mu_1 + \lambda_0$ összeadás valójában olyan kivonás, amelynek során értékes számjegyeket veszíthetünk el: $\lambda_0 \approx \lambda_n \gg \lambda_1$, és

$$\mu_1 + \lambda_0 \approx (\lambda_1 - \lambda_n) + \lambda_n = \lambda_n - (\lambda_n - \lambda_1).$$

Általános valós mátrix esetén is, ha teljesül a (2.2) dominanciafeltétel, érdemes a $\lambda_0 = \|\mathbf{A}\|$ eltolással próbálkozni. A domináns sajátérték lehet negatív is, ekkor az eltolás

ezen nem változtat, így a változatlan számítási eredményből ezt észrevehetjük. Ilyenkor pedig következőnek $\lambda_0 = -\|\mathbf{A}\|$ -val lehet próbálkozni.

Más eltolást is lehet alkalmazni. Például ha sejtésünk szerint a konvergencia azért nem teljesül, mert van egy komplex sajátértékpár, akkor javasolt egy tisztán képzetes eltolást használni. Ekkor a két konjugált sajátérték közül már csak az egyik lesz abszolút értékben maximális.

2.2. Inverz iteráció

Az inverz iteráció elvileg úgy jön létre, hogy a hatványmódszert az inverz mátrixra alkalmazzuk. De a korábbi (2.1) reláció helyett most nem az $\mathbf{y}^m := \mathbf{A}^{-1}\mathbf{y}^{m-1}$ műveletet fogjuk végrehajtani, hanem az $\mathbf{A}\mathbf{y}^m = \mathbf{y}^{m-1}$ lineáris rendszert oldjuk meg LU -felbontással. Ehhez nyilván az iteráció előtt LU -felbontást készítünk \mathbf{A} -ból, majd szokott módon dolgozunk, megoldva az m -ciklusban változatlan L -lel és U -val az

$$L\mathbf{z} = \mathbf{y}^{m-1}, \quad \text{és} \quad U\mathbf{y}^m = \mathbf{z}$$

rendszereket, ahol \mathbf{z} segédvektor. Itt is érdemes a kapott vektorokat normalizálni.

2.2.1. Konvergenciafeltételek

A hatványmódszer konvergenciafeltételeinek megfelelői itt a következők:

Legyen $\mathbf{x}^0 \neq \mathbf{0}$ egy kezdeti vektor és készüljön az $\{\mathbf{x}^m\}$ sorozat az

$$m = 1, 2, \dots : L\mathbf{z} = \mathbf{x}^{m-1}, \quad U\mathbf{y} = \mathbf{z}, \quad \mathbf{x}^m = \frac{\mathbf{y}}{\|\mathbf{y}\|}$$

iterációval. Ekkor, ha \mathbf{A} normális és érvényes

$$(\mathbf{x}^0, \mathbf{v}^{(1)}) \neq 0, \quad |\lambda_1| < |\lambda_2| \leq \dots \leq |\lambda_n|,$$

akkor a sorozat konvergál $\mathbf{v}^{(1)}$ -hez (vagy egy többszöröséhez), a legkisebb sajátértékhez tartozó sajátvektorhoz.

Ez teljesül, hiszen ha \mathbf{A} reguláris és az (2.7) egyenletet megszorozzuk \mathbf{A}^{-1} -gyel, akkor azt kapjuk, hogy $\mathbf{x} = \lambda\mathbf{A}^{-1}\mathbf{x}$. 0 nem sajátértéke \mathbf{A} , mivel az reguláris, így ezt a relációt λ -val eloszthatjuk:

$$\frac{1}{\lambda}\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}, \quad \mathbf{x} \neq \mathbf{0}.$$

Vagyis az inverz mátrix $\lambda(\mathbf{A}^{-1})$ sajátértékei az eredeti mátrix $\lambda(\mathbf{A})$ sajátértékeinek reciprokai, azaz $\lambda(\mathbf{A}^{-1}) = \frac{1}{\lambda(\mathbf{A})}$, a sajátvektorok ugyanazok. És ekkor ha $\lambda_1(\mathbf{A})$ a legkisebb abszolút értékű sajátértéke \mathbf{A} -nak, akkor $\frac{1}{\lambda_1(\mathbf{A})}$ a maximális abszolút értékű sajátértéke \mathbf{A}^{-1} -nek. Alkalmazva az inverz iterációt, megkapjuk a hozzátartozó $\mathbf{v}^{(1)}$ sajátvektor közelítését, és ezután ennek Rayleigh-hányadosából a $\lambda_1(\mathbf{A})$ sajátérték közelítését.

Az inverz iteráció azért hasznos, mert nemcsak maximális és minimális, hanem közbülső sajátérték meghatározására is alkalmazható, ha használjuk az eltolást is. Ez nagy előny a hatványmódszerrel szemben.

Eltolás esetén a módszer ahhoz a sajátvektorhoz konvergál, amelynek a hozzá tartozó sajátértéke a legközelebb van a λ_0 eltoláshoz, hiszen ha

$$\min_i |\lambda_i - \lambda_0| =: |\lambda_{i_0} - \lambda_0| > 0,$$

akkor $\frac{1}{\lambda_{i_0} - \lambda_0}$ lesz $(\mathbf{A} - \lambda_0 \mathbf{I})^{-1}$ -nek abszolút értékben a legnagyobb sajátértéke.

A konvergencia sebessége most a

$$\max_{i \neq i_0} \frac{|\lambda_{i_0} - \lambda_0|}{|\lambda_i - \lambda_0|}$$

mennyiségtől függ, amely különösen akkor lesz közel nullához, ha λ_0 már jó közelítése a keresett λ_{i_0} sajátértéknek.

Itt is érdemes a kapott eredménnyel a programot újra indítani úgy, hogy λ_0 eltolás gyanánt a sajátérték éppen kiszámított közelítésére térünk át.

Ha a λ_0 eltolás túl közel van egy sajátértékhez, akkor ugyan nagy hibával kapjuk az $(\mathbf{A} - \lambda_0 \mathbf{I})\mathbf{x}^m = \mathbf{y}^{m-1}$ egyenlet megoldását, de a hiba majdnem teljesen a közeli sajátértékhez tartozó sajátvektor irányában keletkezik, és ez jó nekünk.

Ha a λ_0 eltolás egyenlő egy sajátértékkel, akkor az LU-felbontás során észrevesszük a szingularitást, és ekkor az eredeti egyenletrendszert nem tudjuk megoldani, de az $(\mathbf{A} - \lambda_0 \mathbf{I})\mathbf{x}^m = \mathbf{0}$ rendszerhez találhatunk nemtriviális megoldást. Ez lesz a keresett sajátvektor.

Fontos, hogy az inverz iteráció használatakor is lehetséges olyan eset, amikor két legközelebbi sajátérték létezik a λ_0 eltoláshoz. Ilyenkor ugyanazok a problémák lépnek fel, mint a hatványmódszer esetén. Ezért itt is célszerű, hogy az eljárás végén ellenőrizzük, vajon sajátvektorhoz történt-e a konvergencia.

Az inverz iteráció algoritmusát a hatványmódszeréből könnyedén meg lehet kapni néhány kisebb átalakítással, így ezt külön most nem ismertetem.

2.3. Householder-eljárás

Legyen szimmetrikus és ortogonális a következő mátrix (neve: Householder-mátrix):

$$\mathbf{Q} = \mathbf{I} - c\mathbf{u}\mathbf{u}^T, \quad c = \frac{2}{\|\mathbf{u}\|^2}. \quad (2.8)$$

A sajátérték megtartása érdekében a következő transzformációt fogjuk alkalmazni:

$$\mathbf{A} \rightarrow \mathbf{Q}\mathbf{A}\mathbf{Q} =: \mathbf{A}_1. \quad (2.9)$$

Ha ezután \mathbf{v} az \mathbf{A}_1 mátrix sajátvektora, akkor $\mathbf{Q}\mathbf{v}$ az \mathbf{A} mátrixé.

Az \mathbf{u} vektort úgy választjuk, hogy \mathbf{A}_1 első oszlopának minden eleme nulla legyen, kivéve az első két elemét. Ezután \mathbf{A}_1 -ből az első sort és oszlopot tovább nem vesszük figyelembe és analóg transzformációt alkalmazunk az $(n-1) \times (n-1)$ -es al mátrixra, és így tovább. Így $n-2$ lépés alatt elérkezünk az \mathbf{A}_{n-2} mátrixhoz, amelynek speciális alakja van: $\mathbf{A}_{n-2} = (b_{ij})$ Hessenberg-alakú lesz: alsó háromszöge – az első mellékátló kivételével – nulla (tehát $b_{ij} = 0$, ha $i > j + 1$). Ha most \mathbf{A} szimmetrikus, akkor \mathbf{A}_{n-2} is szimmetrikus és Hessenberg-alakú, tehát tridiagonális.

Ez a Householder-eljárás; előnye, hogy $O(n^3)$ lépés során, a sajátérték megtartása mellett olyan mátrixot kapunk, amely elemeinek jeletős része nulla. Így még nincsenek meg a sajátértékek, de az ezután bevethető módszerek hatékonyságát növeljük (pl. QR -módszer).

Nézzük most meg az \mathbf{u} vektor választását. Ha $\mathbf{u} = (0, \underline{\mathbf{u}})^T$, akkor

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \underline{\mathbf{Q}} & \\ 0 & & & \end{bmatrix}, \quad \underline{\mathbf{Q}} \in \mathbb{R}^{(n-1) \times (n-1)},$$

Ez a (2.9) transzformáció során éppen jó lesz. Az alábbiakban leírom miért.

A (2.8)-ból kapjuk, hogy $\mathbf{Q}\mathbf{A} = \mathbf{A} - \mathbf{u}\mathbf{v}^T$, ahol $\mathbf{v} := c\mathbf{A}\mathbf{u}$, továbbá

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{A} - \mathbf{v}\mathbf{u}^T - \mathbf{u}\mathbf{v}^T + c\mathbf{u}\mathbf{u}^T\mathbf{v}\mathbf{u}^T = \mathbf{A} - \left(\mathbf{v} - \frac{c}{2}(\mathbf{u}^T\mathbf{v})\mathbf{u}\right)\mathbf{u}^T - \\ &\quad - \mathbf{u}\left(\mathbf{v}^T - \frac{c}{2}(\mathbf{u}^T\mathbf{v})\mathbf{u}^T\right) = \mathbf{A} - \mathbf{w}\mathbf{u}^T - \mathbf{u}\mathbf{w}^T, \end{aligned} \quad (2.10)$$

ahol $\mathbf{w} := \mathbf{v} - \frac{c}{2}(\mathbf{u}^T\mathbf{v})\mathbf{u}$. Legyen most $(\mathbf{B})_1$ valamilyen \mathbf{B} mátrix első oszlopa és $(\mathbf{A})_1 =: \mathbf{a}_1$.

Mivel \mathbf{u} első komponense nulla, a $\mathbf{w}\mathbf{u}^T$ szorzat első oszlopa nulla és így

$$(\mathbf{A}_1)_1 = \mathbf{a}_1 - \mathbf{w}_1\mathbf{u}.$$

Ugyanebből az okból

$$\mathbf{w}_1 = \mathbf{v}_1 - \frac{c}{2}(\mathbf{u}^T\mathbf{v})\mathbf{u}_1 = \mathbf{v}_1,$$

így végül

$$(\mathbf{Q}\mathbf{A}\mathbf{Q})_1 = (\mathbf{A}_1)_1 = \mathbf{a}_1 - \mathbf{v}_1\mathbf{u} = (\mathbf{Q}\mathbf{A})_1.$$

Ez fontos eredmény: $\mathbf{Q}\mathbf{A}\mathbf{Q}$ első oszlopa megegyezik $\mathbf{Q}\mathbf{A}$ első oszlopával, és így az \mathbf{u} vektort választhatjuk a következőképpen:

$$\mathbf{u}_1 = \mathbf{a}_1 + \tau; \quad \tau = \begin{cases} \|\mathbf{a}\|, & \text{ha } \mathbf{a}_1 \geq 0 \\ -\|\mathbf{a}\|, & \text{ha } \mathbf{a}_1 < 0 \end{cases} \quad (2.11)$$

$$\mathbf{u}_i = \mathbf{a}_i, \quad i > 1; \quad \frac{1}{c} = \frac{1}{2}\|\mathbf{u}\|^2 = \tau\mathbf{u}_1.$$

A különbség ehhez képest annyi, hogy már lekötöttük, hogy $\mathbf{u}_1 = 0$. Emiatt csak az \mathbf{u} vektor többi komponensét összefoglaló $\underline{\mathbf{u}}$ választásáról lehet szó (így az eljárás során csak az $i > j + 1$ -pozíciójú elemeit nullázzuk ki).

Az $\underline{\mathbf{u}}$ vektort most (2.11) képletek szerint úgy választjuk, hogy az \mathbf{A} mátrix első oszlopának, vagyis $\mathbf{a}_1 = (a_{11}, \underline{\mathbf{a}}_1)^T$ -nak az \mathbf{a}_1 alsó, $(n-1)$ -dimenziós vektorba elfoglalt része az $\mathbf{e}_1 = (1, 0, 0, \dots, 0)^T$ első $(n-1)$ -dimenziós koordináta egységvektor konstansszorosába menjen át, tehát

$$\underline{\mathbf{Q}}\mathbf{a}_1 = -\tau\underline{\mathbf{e}}_1 \quad \text{és} \quad \mathbf{Q}\mathbf{a}_1 = (a_{11}, -\tau\underline{\mathbf{e}}_1)^T.$$

Ehhez megfelel a

$$\underline{\mathbf{u}} = \underline{\mathbf{a}}_1 + \tau\underline{\mathbf{e}}_1, \quad \tau = s\|\underline{\mathbf{a}}_1\|, \quad s = \begin{cases} +1 & a_{21} \geq 0 \\ -1 & a_{21} < 0 \end{cases}, \quad (2.12)$$

választás, és ekkor

$$c = \frac{2}{\|\mathbf{u}\|^2} = \frac{1}{\tau\mathbf{u}_2}, \quad a_{21}^{(1)} = -\tau.$$

Ezzel megvan az $\mathbf{u} = \mathbf{u}^{(1)}$ vektor.

A második lépésben az $\mathbf{u} = \mathbf{u}^{(2)}$ vektort úgy választjuk, hogy első két komponense nulla legyen, a többi komponens segítségével pedig kinullázzuk \mathbf{A}_1 -ben az $(i, 2)$ -pozíciójú elemeket, $i > 3$, stb., míg nem az $(n-2)$ -edik lépésben a tridiagonális, illetve

Hessenberg-alakú \mathbf{A}_{n-2} mátrixot kapjuk meg olyan $\mathbf{u}^{(n-2)}$ vektor segítségével, amelynek első $n - 2$ komponense mind nulla.

Következzen pár szó az eljárás műveletigényéről. A Householder-transzformáció $\frac{2n^3}{3}$ műveletére itt is szükség van, hiszen az összes \mathbf{QA} mátrixot ki kell számítani, míg a \mathbf{QAQ} mátrixok csak $O(n^2)$ műveletbe kerülnek (itt egy művelet alatt egy összeadást és egy szorzást értünk). A (2.11) képlet alapján, az első lépésben $n(n - 1)$ művelet kell \mathbf{v} előállítására, \mathbf{w} kiszámításának költsége csak $O(n)$, $\mathbf{w}\mathbf{u}^T$ -hez viszont még egyszer $n(n - 1)$ művelet szükséges, ezzel $\mathbf{u}\mathbf{w}^T$ és \mathbf{A}_1 is megvannak. A második lépésben \mathbf{A}_1 első sora és oszlópa már nem érdekel, így lényegében $2(n - 1)(n - 2)$ művelet szükséges, stb., az $(n - 2)$ -edik lépés után összesen $\frac{2n^3}{3} + O(n^2)$ művelet.

Egyébként $\mathbf{u}^T \mathbf{w} = \mathbf{u}^T \mathbf{v} (1 - \frac{c\|\mathbf{u}\|^2}{2}) = 0$, c meghatározása szerint. Így a Householder-eljárás végén érvényesek az

$$\mathbf{A} = \mathbf{A}_{n-2} - \sum_{i=1}^{n-2} \left(\mathbf{w}^i (\mathbf{u}^i)^T + \mathbf{u}^i (\mathbf{w}^i)^T \right), \quad (\mathbf{u}^i)^T \mathbf{w}^i = 0,$$

összefüggések.

2.3.1. Megjegyzés. Ha \mathbf{A} ritkamátrix, a tridiagonális, illetve Hessenberg-alakra hozást célszerűbb forgatási (és nem Householder) mátrixokkal végezni.

3. fejezet

A sajátértékeket egyszerre közelítő eljárások

Most következzenek a sajátértékeket egyszerre közelítő módszerek. A fejezetet [1] alapján készítettem.

3.1. A Jacobi-módszer

Nézzük először a Jacobi-iterációt, amely ugyan egyszerű, mégis 1846 óta az egyik legolcsóbb eljárás kis méretű ($n < 10$) szimmetrikus valós mátrixok sajátértékeinek és sajátvektorainak meghatározására. A módszer célja a mátrix főátlón kívüli elemeinek kinullázása iteratív eljárással úgy, hogy végül eljussunk a $\text{diag}(\lambda_i(\mathbf{A}))$ mátrixhoz.

3.1.1. Leírás és konvergencia

Legyen $\mathbf{A} = \mathbf{A}^T = (a_{ij})$ és

$$a_{kl} := \max_{i \neq j} |a_{ij}| > 0, \quad k < l.$$

Ekkor úgy választjuk a $\mathbf{Q}_{kl} = \mathbf{Q}_{kl}(\theta)$ forgatási mátrixot, hogy az

$$\mathbf{A}_1 := \mathbf{Q}_{kl}^{-1} \mathbf{A} \mathbf{Q}_{kl} = (a_{ij}^{(1)}) \tag{3.1}$$

mátrixban $a_{kl}^{(1)} = 0$ lesz. A \mathbf{Q}_{kl} mátrix ortogonális, ezért $\mathbf{Q}_{kl}^{-1} = \mathbf{Q}_{kl}^T$, és ezért \mathbf{A} -val együtt \mathbf{A}_1 is szimmetrikus lesz. Továbbá, a (3.1) transzformáció változatlanul hagyja a sajátértéket.

Foglalkozzunk most az $a_{kl}^{(1)}$ elem kiszámításával. Szorozzuk össze a \mathbf{Q}_{kl}^T , \mathbf{A} és \mathbf{Q}_{kl} mátrixnak a (k, k) , (k, l) , (l, k) , (l, l) pozíciójú elemeiből összeállított 2×2 -es almatríxokat. Ekkor azt kapjuk, hogy

$$a_{kl}^{(1)} = \cos \theta (a_{kk} \sin \theta + a_{kl} \cos \theta) - \sin \theta (a_{lk} \sin \theta + a_{ll} \cos \theta).$$

Ez akkor nulla, amikor

$$0 = \cos \theta \sin \theta (a_{kk} - a_{ll}) + (\cos^2 \theta - \sin^2 \theta) a_{kl},$$

vagyis ha

$$\cot 2\theta = \frac{a_{kk} - a_{ll}}{2a_{kl}}.$$

Innen $\sin \theta$ és $\cos \theta$ értékei megfelelő trigonometrikus képletek segítségével kiszámíthatók.

\mathbf{A} és \mathbf{A}_1 elemei csak $i = k, l$ és $j = k, l$ -re különböznek, vagyis csak a k -edik és l -edik sor, illetve oszlop változik. Ekkor érvényesek a következő formulák:

$$\begin{aligned} a_{ki}^{(1)} &= a_{ik}^{(1)} = a_{ik} \cos \theta - a_{il} \sin \theta, \quad i \neq k, l \\ a_{il}^{(1)} &= a_{li}^{(1)} = a_{ik} \sin \theta + a_{il} \cos \theta, \quad i \neq k, l \\ a_{kl}^{(1)} &= 0, \quad a_{kk}^{(1)} = a_{kk} \cos^2 \theta - 2a_{kl} \cos \theta \sin \theta + a_{ll} \sin^2 \theta, \\ a_{lk}^{(1)} &= 0, \quad a_{ll}^{(1)} = a_{kk} \sin^2 \theta + 2a_{kl} \cos \theta \sin \theta + a_{ll} \cos^2 \theta. \end{aligned} \tag{3.2}$$

Hasonló módon, mint ahogyan $\mathbf{A} = \mathbf{A}_0$ -ból \mathbf{A}_1 -et, kapjuk az \mathbf{A}_2 , \mathbf{A}_3, \dots mátrixokat is (mindig egy főátlón kívüli elemet kinullázva), és ezen mátrixok sajátértékei mind megegyeznek. Könnyen megeshet, hogy egy kinullázott elem újra megjelenik, ezért ez egy iteratív folyamat lesz, nem pedig véges.

Legyen az m -edik transzformáció eredménye az $\mathbf{A}_m = \text{diag}(a_{ii}^{(m)}) + \mathbf{B}_m$ mátrix. Ekkor a \mathbf{B}_m mátrix főátlóján mind nullák állnak. Mutassuk meg, hogy

$$\|\mathbf{B}_m\|_F := \left(\sum_{i,j} \left(b_{ij}^{(m)} \right)^2 \right)^{\frac{1}{2}} \rightarrow 0, \quad m \rightarrow \infty.$$

Itt $\|\cdot\|_F$ a Frobenius-normát jelöli. \mathbf{B}_m definíciója szerint

$$\|\mathbf{B}_m\|_F^2 = \sum_{i \neq j} \left(a_{ij}^{(m)} \right)^2 = \sum_{\substack{i \neq j \\ i, j \neq k, l}} \left(a_{ij}^{(m-1)} \right)^2 + 2 \cdot \sum_{i \neq k, l} \left[\left(a_{ik}^{(m)} \right)^2 + \left(a_{il}^{(m)} \right)^2 \right],$$

ahol figyelembe vettük, hogy $a_{kl}^{(m)} = a_{lk}^{(m)} = 0$.

Viszont

$$\left(a_{ik}^{(m)}\right)^2 + \left(a_{il}^{(m)}\right)^2 = \left(a_{ik}^{(m-1)}\right)^2 + \left(a_{il}^{(m-1)}\right)^2, \quad i \neq k, l.$$

Ezzel együtt

$$\|\mathbf{B}_m\|_F^2 = \|\mathbf{B}_{m-1}\|_F^2 - 2\left(a_{kl}^{(m-1)}\right)^2.$$

A (k, l) indexpár kiválasztása szerint

$$\|\mathbf{B}_{m-1}\|_F^2 \leq n(n-1) \max_{i \neq j} \left(a_{ij}^{(m-1)}\right)^2 = n(n-1) \left(a_{kl}^{(m-1)}\right)^2,$$

tehát

$$\|\mathbf{B}_m\|_F^2 \leq \|\mathbf{B}_{m-1}\|_F^2 \left(1 - \frac{2}{n(n-1)}\right) \leq \left(1 - \frac{2}{n(n-1)}\right)^m \|\mathbf{B}_0\|_F^2. \quad (3.3)$$

3.1.1. Tétel. *A Jacobi-módszer szimmetrikus mátrix esetén konvergens.*

Bizonyítás. A (3.3) becslés szerint nagy m -re \mathbf{A}_m főátlón kívüli elemei kicsik. Legyen például

$$\sum_{i \neq j} \left|a_{ij}^{(m)}\right| < \varepsilon \quad (3.4)$$

minden i -re. Ekkor Gersgorin-tétele szerint az összes sajátérték az

$$\left|a_{ii}^{(m)} - \lambda\right| < \varepsilon$$

körökben található. \square

3.1.2. Megjegyzés. A Jacobi-módszer akkor is konvergál, ha a mátrix normális vagy ha nemszimmetrikus, de minden sajátértéke valós.

3.1.2. Gyakorlati szempontok

A Jacobi-módszer során ciklusosan járunk el, mivel a fenti módszer túl sok időt vesz igénybe, pusztán a maximális főátlón kívüli elem kiválasztásához is $O(n^2)$ összehasonlítás szükséges. A ciklus során az egyes főátlón kívüli elemeknek megfelelő \mathbf{Q}_{kl} ortogonális mátrixokat alkalmazzuk rögzített sorrendben, például a felső háromszög elemeit átlónként végigjárva. Az első teljes ciklusban küszöbértékeket adunk meg a gyorsítás érdekében: amely kinullázandó elem kisebb ennél az értéknél, azt kihagyjuk.

Ha nincsenek abszolút értékben megegyező sajátértékek és a (3.3)-ban szereplő ε elég kicsi, akkor a módszer négyzetesen konvergál. Ha $n\varepsilon < 1$ és az \mathbf{A}_m főátlóbeli

elemek távolsága egymástól nagy ε -hoz képest, akkor egy teljes ciklus után (3.4)-ben ε helyett már $(\varepsilon n)^2$ áll.

A tapasztalatok szerint legfeljebb öt teljes ciklus kell ahhoz, hogy az összes sajátértéket kiszámítsuk maximális pontossággal. Ekkor a Jacobi-módszer műveletigénye durván $10n^3$, mivel a (3.2) képletekből látszik, hogy lényegében $4n$ szorzásra és $2n$ összeadásra van szükség egy $\mathbf{A}_m \rightarrow \mathbf{A}_{m+1}$ lépés végrehajtásához, a főátlón kívüli elemek száma pedig $\frac{n(n-1)}{2}$.

Következzen a $\sin \theta$ és $\cos \theta$ értékek kiszámolása (maga a θ forgatási szög nem fontos).

Használjuk a következő képleteket, az adott $p := \cot 2\theta$ -ból kiindulva:

$$\cos 2\theta = \frac{p}{(1+p^2)^{\frac{1}{2}}}, \quad \cos \theta = \left(\frac{1 + \cos 2\theta}{2} \right)^{\frac{1}{2}}, \quad \sin \theta = \left(\frac{1 - \cos 2\theta}{2} \right)^{\frac{1}{2}}.$$

A Jacobi-módszer előrehaladásával legtöbbször $p = \cot 2\theta \rightarrow \infty$, $\cos 2\theta \rightarrow 1$, $\cos \theta \rightarrow 1$, $\sin \theta \rightarrow 0$. Amikor azonban $p \gg 1$ (pontosabban $p > 1/\varepsilon_1^{\frac{1}{2}}$), akkor $\text{fl}(p^2 + 1) = p^2$, így $\cos 2\theta = 1$, $\cos \theta = 1$, és $\sin \theta = 0$. Ekkor megáll az iteráció, mivel $a_{kl}^{(m+1)} = a_{kl}^{(m)}$, és nem $a_{kl}^{(m+1)} = 0$.

Használjunk más képleteket is: Legyen $t := \tan \theta$. Ekkor $t^2 + 2pt - 1 = 0$. Innen

$$t = \tan \theta = \begin{cases} 1, & p = 0 \\ \frac{1}{p + \text{sgn } p(1+p^2)^{\frac{1}{2}}}, & p \neq 0 \end{cases},$$

$$\cos \theta = \frac{1}{(1+t^2)^{\frac{1}{2}}}, \quad \sin \theta = t \cos \theta.$$

Ekkor ha p nagy, akkor t kicsi és ha $t < \varepsilon_1^{\frac{1}{2}}$, akkor $\text{fl}(1+t^2) = 1$, tehát $\cos \theta = 1$, de $\sin \theta = t$ és nem nulla. (Ekkor nem $\cos^2 \theta + \sin^2 \theta = 1$, hanem $\text{fl}(\cos^2 \theta + \sin^2 \theta) = 1$.)

A sajátvektorok a (3.1) transzformáció alatt a következőképpen változnak: ha \mathbf{v} az \mathbf{A}_1 sajátvektora, akkor $\mathbf{Q}_{kl}\mathbf{v}$ az \mathbf{A} mátrixé, ezért fontos, hogy a \mathbf{Q}_{kl} mátrixok szorzatát folyamatosan figyeljük. Legyen

$$\mathbf{Q}_0 := \mathbf{I}, \quad \mathbf{Q}_m = \mathbf{Q}_{kl}^{(m)} \mathbf{Q}_{m-1}, \quad m = 1, 2, \dots, \quad (3.5)$$

ahol k és l indexek m -mel együtt változnak.

Ha az \mathbf{A}_m mátrix diagonális, akkor ennek j -edik sajátvektora e^j , az \mathbf{A} mátrix (közelítő) j -edik sajátvektora $\mathbf{v}^j = \mathbf{Q}_m e^j$.

3.2. QR -módszer

A QR -módszert akkor szoktuk használni, ha sok sajátértéket keresünk.

Tudjuk, hogy tetszőleges mátrixot, így az $\mathbf{A}_0 := \mathbf{A}$ mátrixot is, \mathbf{Q}_0 ortogonális és \mathbf{R}_0 felső háromszög mátrixra bonthatjuk fel (például Householder-transzformációval):

$$\mathbf{A} = \mathbf{A}_0 = \mathbf{Q}_0 \mathbf{R}_0$$

Ezután képezzük az

$$\mathbf{A}_1 := \mathbf{R}_0 \mathbf{Q}_0 = \mathbf{Q}_0^T \mathbf{Q}_0 \mathbf{R}_0 \mathbf{Q}_0 = \mathbf{Q}_0^T \mathbf{A}_0 \mathbf{Q}_0 \quad (3.6)$$

mátrixot. (3.6)-ból látjuk, hogy \mathbf{A}_0 és \mathbf{A}_1 sajátértékei megegyeznek. (Ha \mathbf{v} az \mathbf{A}_1 mátrix sajátvektora, akkor $\mathbf{Q}_0 \mathbf{v}$ az \mathbf{A}_0 mátrixé.) Ugyanúgy, mint \mathbf{A}_0 -ból \mathbf{A}_1 -et, kapjuk \mathbf{A}_1 -ből \mathbf{A}_2 -t, és így tovább:

$$\mathbf{A}_0 =: \mathbf{Q}_0 \mathbf{R}_0, \quad \mathbf{A}_m = \mathbf{R}_{m-1} \mathbf{Q}_{m-1}, \quad \mathbf{A}_m = \mathbf{Q}_m \mathbf{R}_m, \quad m = 1, 2, \dots$$

Ez a QR -módszer.

A QR -módszer általa előállított $\{\mathbf{A}_m\}$ sorozat legtöbbször egy diagonális mátrixhoz konvergál, például akkor is, amikor \mathbf{A} szimmetrikus és pozitív definit. Ha az \mathbf{A} telt mátrix, akkor minden QR -lépés $O(n^3)$ műveletet igényel, és a konvergencia is lassú.

A konvergenciát fel lehet gyorsítani, ha \mathbf{A} -t Hessenberg-, illetve tridiagonális alakra hozzuk. Ezek az alakok invariánsak a QR -iteráció alatt, mivel

$$\mathbf{A}_1 = \mathbf{R}_0 \mathbf{Q}_0 = \mathbf{R}_0 \mathbf{Q}_0 \mathbf{R}_0 \mathbf{R}_0^{-1} = \mathbf{R}_0 \mathbf{A}_0 \mathbf{R}_0^{-1},$$

ahol \mathbf{R}_0 és az inverze is felső háromszög: Hessenberg-alakú mátrix szorzata egy felső háromszög mátrixszal Hessenberg-alakú mátrix lesz, valamint \mathbf{A}_1 szimmetrikus, ha \mathbf{A}_0 szimmetrikus volt.

Lássunk még egy azonosságot: az

$$\mathbf{A}_{m+1} = \mathbf{Q}_m^T \mathbf{A}_m \mathbf{Q}_m = \mathbf{Q}_m^T \mathbf{Q}_{m-1}^T \mathbf{A}_{m-1} \mathbf{Q}_{m-1} \mathbf{Q}_m = \dots = \mathbf{Q}_m^T \dots \mathbf{Q}_0^T \mathbf{A}_0 \mathbf{Q}_0 \dots \mathbf{Q}_m,$$

egyenlőségekből a

$$\mathbf{P}_m := \mathbf{Q}_0 \dots \mathbf{Q}_m$$

jelöléssel azt kapjuk, hogy

$$\mathbf{P}_m \mathbf{A}_{m+1} = \mathbf{A} \mathbf{P}_m \quad (3.7)$$

és

$$\mathbf{P}_m \mathbf{R}_m = \mathbf{P}_{m-1} \mathbf{Q}_m \mathbf{R}_m = \mathbf{P}_{m-1} \mathbf{A}_m = \mathbf{A} \mathbf{P}_{m-1}. \quad (3.8)$$

3.2.1. Tétel. (QR-módszer konvergenciája) Legyen \mathbf{A} szimmetrikus mátrix, $\{\lambda_i\}$ és $\{\mathbf{v}^i\}$ a sajátértékei, illetve sajátvektorai, és teljesüljenek a

$$|\lambda_1| < |\lambda_2| \leq \dots \leq |\lambda_{n-1}| < |\lambda_n|,$$

$$(\mathbf{e}_1, \mathbf{v}^n) \neq 0, \quad (\mathbf{e}_n, \mathbf{v}^1) \neq 0,$$

összefüggések. Ekkor a QR-módszer által előállított $\{\mathbf{A}_m\}$ mátrixsorozatra érvényes vagy az, hogy $m \rightarrow \infty$ és

$$\mathbf{A}_m \mathbf{e}_1 \rightarrow \lambda_1 \mathbf{e}_1, \quad \mathbf{A}_m \mathbf{e}_n \rightarrow \lambda_n \mathbf{e}_n, \quad \text{azaz} \quad a_{11}^{(m)} \rightarrow \lambda_1, \quad a_{nn}^{(m)} \rightarrow \lambda_n,$$

vagy pedig, véges m -nél, (nulla) sajátértéket ér el.

Bizonyítás. Szorozzuk meg a (3.8)-at az első egységvektorral. Ekkor

$$r_{11}^{(m)} \mathbf{P}_m \mathbf{e}_1 = \mathbf{A} \mathbf{P}_{m-1} \mathbf{e}_1,$$

mivel \mathbf{R}_m felső háromszög mátrix. Legyen

$$\mathbf{y}_0 := \mathbf{e}_1, \quad \mathbf{y}_m := \mathbf{P}_m \mathbf{e}_1,$$

ekkor $\|\mathbf{y}_m\| = \|\mathbf{P}_m \mathbf{e}_1\|$, ugyanis $\mathbf{P}_m = \mathbf{Q}_0 \dots \mathbf{Q}_m$ ortogonális. Ezért, ha $r_{11}^{(m)} \neq 0$ minden m -re, akkor teljesülnek a hatványmódszer feltételei, így

$$\mathbf{y}_m = \left(r_{11}^{(m)}\right)^{-1} \mathbf{A} \mathbf{y}_{m-1} \rightarrow \mathbf{v}^n,$$

. Ezért

$$0 \leftarrow \mathbf{A} \mathbf{y}_m - \lambda_n \mathbf{y}_m = \mathbf{A} \mathbf{P}_m \mathbf{e}_1 - \lambda_n \mathbf{P}_m \mathbf{e}_1 = \mathbf{P}_m (\mathbf{A}_{m+1} \mathbf{e}_1 - \lambda_n \mathbf{e}_1),$$

ahol használtuk (3.7)-et is. Tehát $\mathbf{A}_{m+1} \mathbf{e}_1 \rightarrow \lambda_n \mathbf{e}_1$.

(Ha $r_{11}^{(m)} = 0$ valamilyen m -re, akkor $0 = \mathbf{A} \mathbf{P}_{m-1} \mathbf{e}_1 = \mathbf{A} \mathbf{y}_{m-1}$, vagyis \mathbf{y}_{m-1} sajátvektor és 0 a sajátérték. Ekkor \mathbf{A}_m első oszlopa és első sora is csupa nulla.)

Következzen az állítás második, λ_1 -re vonatkozó fele. (3.8)-ból transzponálással kapjuk, hogy

$$\mathbf{R}_m^T \mathbf{P}_m^T = \mathbf{P}_{m-1}^T \mathbf{A}, \quad \text{tehát} \quad \mathbf{P}_{m-1} \mathbf{R}_m^T = \mathbf{A} \mathbf{P}_m.$$

Sorozzuk meg ezt \mathbf{e}_n -nel. \mathbf{R}_m^T alsó háromszög mátrix, ezért a $\mathbf{z}_m := \mathbf{P}_m \mathbf{e}_n$ -re azt kapjuk, hogy

$$\|\mathbf{z}_m\| = 1, \quad \mathbf{A} \mathbf{z}_m = r_{nn}^{(m)} \mathbf{z}_{m-1}.$$

Ekkor ha minden m -re fennáll, hogy $r_{nn}^{(m)} \neq 0$, akkor teljesülnek az inverz iteráció feltételei, így $\mathbf{z}_m \rightarrow \mathbf{v}^1$, $\mathbf{A}_m \mathbf{e}_n \rightarrow \lambda_1 \mathbf{e}_n$. Ha pedig valamelyik m -re $r_{nn}^{(m)} = 0$, akkor megint 0 a sajátérték és \mathbf{z}_m a sajátvektor. \square

Legyen most \mathbf{A} szimmetrikus és tridiagonális, tehát minden m -re

$$\mathbf{A}_m = \text{tridiag}\left(b_{i-1}^{(m)}, a_i^{(m)}, b_i^{(m)}\right),$$

és legyen \mathbf{B}_m az \mathbf{A}_m mátrix alsó jobb 2×2 -es almátrixa,

$$\mathbf{B}_m := \begin{bmatrix} a_{n-1}^{(m)} & b_{n-1}^{(m)} \\ b_{n-1}^{(m)} & a_n^{(m)} \end{bmatrix}.$$

Figyeljük meg az $a_n^{(m)}$ elem közelében lévő $b_{n-1}^{(m)}$ és $b_{n-2}^{(m)}$ elemeket. Amikor $b_{n-1}^{(m)}$ elegendően kicsi, konkrétan

$$\left|b_{n-1}^{(m)}\right| \leq \varepsilon \left(\left|a_n^{(m)}\right| + \left|a_{n-1}^{(m)}\right|\right),$$

akkor (a Gersgorin-tétel alapján) $a_n^{(m)}$ lesz a sajátérték közelítése. Ekkor az $n \rightarrow n-1$ behelyettesítés után továbblépünk. (Fontos megemlíteni a 3.2.1 tétel bizonyításában előforduló egyik speciális esetet: ha ott $r_{nn}^{(m)} = 0$ volt, akkor $b_{n-1}^{(m+1)} = a_n^{(m+1)} = 0$.) Amennyiben $b_{n-1}^{(m)}$ nem csökken (és ilyen előfordulhat), de $b_{n-2}^{(m)}$ elég kicsi, \mathbf{B}_m sajátértékeit mint \mathbf{A} közelítő sajátértékeit vesszük, és az $n \rightarrow n-2$ helyettesítés után tovább iterálunk.

A konvergenciát tovább lehet gyorsítani, ha bevezetünk egy σ_m eltolás paramétert. Ekkor legyen

$$\mathbf{A}_0 = \mathbf{A},$$

$$\mathbf{Q}_m^T (\mathbf{A}_m - \sigma_m \mathbf{I}) = \mathbf{R}_m,$$

$$\mathbf{R}_m \mathbf{Q}_m + \sigma_m \mathbf{I} = \mathbf{A}_{m+1}, \quad m = 1, 2, \dots$$

Ez az eltolásos QR -módszer. Ilyenkor is szimmetrikus és tridiagonális lesz \mathbf{A} -val együtt minden \mathbf{A}_m mátrix. A fent említett \mathbf{B}_m mátrixnak azt a sajátértékét vesszük σ_m paraméternek, amely közelebb áll $a_n^{(m)}$ -hez.

Az eltolásos QR -módszer konvergenciája köbös: ha \mathbf{A}_m -ben ε a sajátérték hibája, \mathbf{A}_{m+1} -ben az már $O(\varepsilon^3)$. Átlagban kevesebb, mint 2 iteráció alatt kerül kiszámításra egy sajátérték.

Ha a mátrix nem szimmetrikus, akkor először felső Hessenberg-alakra kell hoznunk, és ez $\frac{5n^3}{3}$ műveletbe kerül. Az iteráció alatt figyelni kell arra, hogy \mathbf{A} -nak konjugált

komplex sajátértékpárjai vannak. A QR -iteráció ekkor egy általánosított Schur-alakra transzformálja a mátrixot, ami azt jelenti, hogy a főátlón lehetnek 2×2 -es blokkmátrixok, amelyek a komplex sajátértékeket tartalmazzák. Ekkor a konvergencia négyzetes.

A QR -módszert akkor ajánlott alkalmazni, ha az n sajátérték legalább negyedét keressük, az eltolásos QR -módszert pedig, akkor, ha minden sajátértéket ki kell számolnunk, ez ugyanis nagyjából tízszer gyorsabb a Jacobi-iterációnál.

Tehát a QR -módszer hasznos, ha sok sajátérték kell, azonban ha például egy telt mátrixnak csak a legkisebb sajátértéke és a hozzá tartozó sajátvektora keresett, akkor jobban járunk az inverz iterációval. Ez látszik abból, hogy az inverz iterációval négy lépést tehetünk, azalatt míg a QR -módszer elvégzi a szükséges Householder-eljárást és megcsinálja a \mathbf{Q} mátrix explicit képzését (mindkettő $\frac{2n^3}{3}$ műveletet igényel). A négy lépés pedig az inverz iterációval legtöbbször elég is a keresett sajátérték és sajátvektor meghatározásához.

4. fejezet

Megvalósítás MATLAB-ban, tesztfeladatok

A következőkben ismertetem néhány numerikus módszer MATLAB-ban megírt kódját, majd mutatok pár érdekes tesztfeladatot. Mielőtt azonban ennek nekikezdek, leírom, hogyan tudjuk a MATLAB segítségével rögtön megadni a sajátértékeket és sajátvektorokat. Ehhez a [2] művet használtam, a kódokhoz pedig a [5, 6, 7] honlapokat.

A MATLAB programcsomag segítségével a teljes standard sajátérték-feladat azonnal megoldható, vagyis adott $n \times n$ -es \mathbf{A} mátrixhoz meghatározható az összes nemnulla \mathbf{x} vektor, amelyre (alkalmas λ komplex számmal, azaz a sajátértékkel) teljesül az alábbi egyenlet:

$$\mathbf{Ax} = \lambda\mathbf{x}.$$

A sajátértékekből pontosan n darab van, és a különböző sajátértékekhez legalább egy sajátvektor tartozik. Két utasítás van, mellyel könnyedén elérhetjük a standard sajátérték-feladat megoldásait:

```
d=eig(A);      % Ekkor csak a sajátértékeket kapjuk  
[V,D]=eig(A); % V a sajátvektorokból áll, D a sajátértékekből
```

Az első utasítás eredménye a \mathbf{d} oszlopvektor, amelynek komponensei az \mathbf{A} sajátértékei.

Az `eig(A)` akkor is működik, ha a mátrix komplex. Ha valós, attól a sajátértékei még lehetnek komplexek, még akkor is, ha az \mathbf{A} mátrix meggyőződésünk szerint szimmetrikus, de kerekítési hibák miatt nem az. Ekkor általában a sajátértékek képzetes részei abszolút értékben kicsik. Ilyen esetben a `d=real(d)` utasítással egyszerűen nullává tesszük a képzetes részt.

Más lehetőség, hogy szimmetrizálunk és csak utána használjuk az `eig` utasítást, a szimmetrizálás időigényes lehet.

A második, sajátvektorokat is kiszámító utasítás csak akkor működik, ha \mathbf{A} telt mátrix. Az utasítás eredménye a két $n \times n$ -méretű \mathbf{V}, \mathbf{D} mátrix. Ezek közül \mathbf{V} oszlopai az \mathbf{A} mátrix sajátvektorait adják, \mathbf{D} pedig egy diagonálmátrix, főátlójában a sajátértékekkel.

4.1. Módszerek kódja

Hatványmódszer

```
function [domsajatertek,v,hiba,iteracio] = hatvmodszer(A,x,epsz,maxit)

hiba=inf;
iteracio=0;
while hiba >= epsz & iteracio < maxit
    z=x/norm(x,2);
    x=A*z;
    lambda1=z'*x;
    if iteracio > 0,
        hiba=abs( lambda1-lambda0 )/abs( lambda0 );
    end
    lambda0=lambda1;
    iteracio=iteracio+1;
end
domsajatertek=lambda1;
v=z;
```

Inverz iteráció

```
function [domsajatertek,v,hiba,iteracio] = inviteracio(A,x,epsz,maxit)

hiba=inf;
iteracio=0;
while hiba >= epsz & iteracio < maxit
```

```

z=x/norm(x,2);
x=A\z;
lambda1=z'*x;
if iteracio > 0,
    hiba=abs( lambda1-alpha0 )/abs( lambda0 );
end
lambda0=lambda1;
iteracio=iteracio+1;
end
domsajatertek=1/lambda1;
v=z;

```

Mielőtt továbbhaladunk talán érdemes megjegyezni, hogy a két kód mennyire hasonlít egymásra. Gyakorlatilag néhány apró eltérést leszámítva teljesen megegyeznek. Ez nyilván abból adódik, hogy az inverz iterációt a hatványmódszerből kapjuk, kis átalakítással. Itt jól látható, hogy kis munkával azonnal megkaptuk az abszolút értékben legkisebb, illetve legnagyobb sajátértéket.

Jacobi-iteráció

A forgatást végző program, amely A (i, j) , illetve (j, i) pozíciójú elemeit kinullázza:

```

function [B,Q] = forgat(A,i,j)

n=max(size(A));
Q=eye(n,n);
p=(A(j,j)-A(i,i))/2/A(i,j);
c2=p/sqrt(1+p^2);
c=sqrt((1+c2)/2);
s=sqrt((1-c2)/2);
Q(i,i)=c;    Q(i,j)=s;    Q(j,i)=-s;    Q(j,j)=c;
B=Q'*A*Q;
B(i,j)=0; B(j,i)=0;

```

A ciklikus Jacobi-módszer a mátrix főátlón kívüli ϵ -nál kisebb elemein megy végig. A `maxiteracio` paraméterében adjuk meg a ciklusok számát. Az ϵ és a `maxiteracio` paraméterek értékeit egymáshoz igazítva állítsuk be. A kinullázott elemek a következő

lépés után már nem lesznek nullák.

A sajátértékek közelítése a \mathbf{B} diagonális elemei lesznek.

```
function [B,Q] = jaccikl(A,epsz,maxiteracio)

n=max(size(A));
Q=eye(n,n);
B=A;
for k=1:maxiteracio
    for i=1:n-1
        for j=(i+1):n
            if abs(B(i,j)) >= epsz
                [SB,SQ]=forgat(B,i,j);    B=SB;    Q=Q*SQ;
            end
        end
    end
end
```

Ennél a megvalósítás már egy fokkal talán nehezebb, hiszen két program kellett, ugyanakkor ez már visszaadja az összes sajátértéket.

4.2. Tesztfeladatok

Lássunk végül néhány tesztfeladatot.

Vegyünk egy olyan mátrixot, amelynek ismerjük a sajátértékeit. Ilyet könnyen gyárthatunk is, hiszen ha van egy \mathbf{A} mátrix, aminek ismerjük a sajátértékeit – mivel például diagonális és így könnyen leolvashatjuk róla azokat – és ezt megszorozzuk balról egy \mathbf{B} tetszőleges mátrixszal, jobbról pedig ennek a \mathbf{B} mátrixnak az inverzével, akkor egy olyan mátrixot kapunk, melynek sajátértékei megegyeznek \mathbf{A} sajátértékeivel, ugyanakkor az új mátrix sajátértékei már nem látszódnak ránézésre. Teszteljük a módszereket egy ilyen mátrixon.

Legyenek például $\mathbf{A}=\text{diag}([1;-1;-0.33;5;3])$, $\mathbf{B}=\text{rand}(5)$ és $\mathbf{S} = \mathbf{BAB}^{-1}$. Keressük \mathbf{S} sajátértékeit. Ekkor a hatványmódszer, illetve az inverz iteráció hibátlanul működik és azt adja, amit várunk. Ekkor természetesen figyelni kell rá, hogy a kezdeti vektor ne legyen merőleges a sajátvektorra, hiszen ilyenkor már rögtön mást kapunk

eredményül, mint amit várnánk. A Jacobi-iterációval azonban messze nem vagyunk sikeresek, de ez sem meglepő, mivel a mátrix, amin végrehajtottuk az iterációt nem volt szimmetrikus. A sebességre egyik esetben sem lehet panaszunk, hiszen a MATLAB segítségével a megírt programok a másodperc töredéke alatt eredményre jutottak. Kisebbségi mátrix esetében pedig még kézzel is jól lehet haladni, hiszen néhány iteráció alatt látható lesz az eredmény.

Vajon tényleg annyira fontos, hogy legyen domináns sajátérték? A válasz annyira nem meglepő módon igen. Legyen most \mathbf{A} a következő mátrix: $\mathbf{A}=\text{diag}([1;-5;-0.33;5;3])$. Jól látható, hogy van kettő olyan sajátértékünk is, amelyeknek abszolút értéke 5. Ekkor ugyan az inverz iteráció még jól működik, de a hatványmódszer már nem, hiszen most még ha nem is szorozzuk be egyéb mátrixszal mint előbb, csak egyszerűen ezen végezzük el az algoritmust, akkor sem adja ki az 5-öt, amit kellene, még magas iterációszám mellett sem. Természetesen ha most úgy vesszük fel a mátrixot, hogy az abszolút értékben legkisebb sajátérték sem lesz egyszeres, akkor az inverz iteráció sem fog már működni.

Végezetül vegyünk egy érdekes mátrixot, az úgynevezett Rosser-mátrixot. Jelöljük ezt \mathbf{R} -rel.

$$\mathbf{R} = \begin{bmatrix} 611 & 196 & -192 & 407 & -2 & -52 & -49 & 29 \\ 196 & 899 & 113 & -192 & -71 & -43 & -8 & -44 \\ -192 & 113 & 899 & 196 & 61 & 49 & 8 & 52 \\ 407 & -192 & 196 & 611 & 8 & 44 & 59 & -23 \\ -8 & -71 & 61 & 8 & 411 & -599 & 208 & 208 \\ -52 & -43 & 49 & 44 & -599 & 411 & 208 & 208 \\ -49 & -8 & 8 & 59 & 208 & 208 & 99 & -911 \\ 29 & -44 & 52 & -23 & 208 & 208 & -911 & 99 \end{bmatrix}$$

Ebben a mátrixban az az érdekes, hogy úgy határozták meg, hogy a sajátértékei között ne legyen domináns, valamint van olyan sajátérték, ami alig tér el egy másiktól, de mégis különböznek és vannak olyanok is, amik teljesen megegyeznek. Ennek a mátrixnak a sajátértékeit nehéz megközelíteni. A fenti eljárások közül a hatványmódszer és az inverz iteráció teljesen cserben hagy minket. A MATLAB beépített sajátérték kereső utasítása ad eredményt, ami jónak tűnik. A Jacobi-iteráció a legeredményesebb a fenti három algoritmus közül, mivel megközelíti a MATLAB által elért eredményeket.

Irodalomjegyzék

- [1] STOYAN GIBERT-TAKÓ GALINA, *Numerikus módszerek 1.*, Typotex, 2005, 188–214, <http://www.tankonyvtar.hu/hu/tartalom/tkt/numerikus-modszerek-1/ch04.html>
- [2] STOYAN GIBERT, *MATLAB*, Typotex, 2005, 67–69
- [3] STOYAN GIBERT, *Numerikus matematika mérnököknek és programozóknak*, Typotex, 2007, 92–104
- [4] FARAGÓ ISTVÁN-HORVÁTH RÓBERT, *Numerikus módszerek*, Typotex, 2013, 20–28, 111–114, <http://tankonyvtar.ttk.bme.hu/pdf/30.pdf>
- [5] <http://www.inf.u-szeged.hu/~berendg/docs/koszi/powerMethod.m>
- [6] <http://www.inf.u-szeged.hu/~berendg/docs/koszi/invPowerMethod.m>
- [7] <http://segabor.web.elte.hu/numanal2/matlab/gepi1.doc>