6. Gram-Schmidt ortogonalizáció, *QR*-felbontás

Az egyszerű lineráris algebrai transzformációk között a harmadik fejezetben megismerkedtünk a vetítő mátrixokkal. E mátrixok alkalmasak arra, hogy a vektorok egy adott készletéből ortogonális vektorokat állítsunk elő. Ha ezen vektorokat egy mátrix oszlopaiba rendezzük, akkor a kapott módszer a mátrix egy újabb felbontását szolgáltatja, ezt hívjuk a QR -felbontásnak.

6.1. A Gram-Schmidt ortogonalizáció

Tegyük fel, van egy lineárisan független vektorokból álló halmaz: $\{a_i\}_{i=1}^k$, $a_i \in \mathbb{R}^m$. Szeretnénk e vektorok felhasználásával olyan ortogonális rendszert készíteni, amellyel a halmaz vektorai előállíthatók. Ekkor eljárhatunk a következőképpen. A készülő ortogonális vektorokat jelöljük q -val.

Az első lépésben válasszuk $q_1 = a_1$-et. A következő vektort készítsük úgy, hogy az a_2 vektort szimmetrikus - vagy más szóhasználattal - ortogonális vetítéssel ortogonalizáljuk q_1-re:

$$\left(I - \frac{q_1 q_1^T}{q_1^T q_1} \right) a_2 = q_2 . \tag{6.1}$$

Beszorzással $q_1^T q_2 = 0$. A következő vektort úgy készítjük, hogy az a_3 vektort q_1 és q_2-re ortogonalizáljuk:

$$\left(I - \frac{q_2 q_2^T}{q_2^T q_2} \right)\left(I - \frac{q_1 q_1^T}{q_1^T q_1} \right) a_3 = q_3 . \tag{6.2}$$

Ismét beszorzással ellenőrízve kapjuk, hogy $q_1 \perp q_3$ és $q_2 \perp q_3$. A továbbiakban vezessük be az i-edik ortogonális vektorhoz a

$$P_i = I - \frac{q_i q_i^T}{q_i^T q_i} \tag{6.3}$$

vetítőmátrixot. Látjuk, ha ezzel a mátrixszal bármely vektorra szorzunk, eredményül a q_i vektorra ortogonális vektorhoz jutunk.

Az $i+1$ -edik ortogonális vektort a következő vetítések sorozatával kapjuk az a_{i+1} vektorból:

$$P_i P_{i-1} \dots P_1 a_{i+1} = q_{i+1} . \tag{6.4}$$

Vegyük észre, a vetítőmátrixok a szorzatban tetszőleges sorrendben írhatók a bennük szereplő vektorok ortogonalitása miatt. Fennáll az összefüggés:

$$P_i P_{i-1} \dots P_1 = \prod_{j=1}^{i} \left(I - \frac{q_j q_j^T}{q_j^T q_j} \right) = I - \sum_{j=1}^{i} \frac{q_j q_j^T}{q_j^T q_j} , \tag{6.5}$$

aminek az igazolását egy feladatra hagyjuk. Ez mutatja, hogy numerikusan kétféle lehetőség van az ortogonalizálásra. Az egyik, amikor a fenti összefüggésben a szummás alakot használjuk. Ekkor minden q_j vektor az a_{i+1} vektorral szorzódik és (6.4), (6.5) összevetéséből kapjuk:

$$q_{i+1} = a_{i+1} - \sum_{j=1}^{i} \frac{q_j q_j^T a_{i+1}}{q_j^T q_j}, \quad \rightarrow \quad a_{i+1} = q_{i+1} + \sum_{j=1}^{i} \frac{q_j q_j^T a_{i+1}}{q_j^T q_j} , \tag{6.6}$$

azaz minden a_{i+1} vektor az ortogonális vektorok segítségével előállítható, ahol a kifejtési együtthatók

$$r_{j,i+1} = \frac{q_j^T a_{i+1}}{q_j^T q_j}. \tag{6.7}$$

Ha (6.4)-ben a mátrixszorzatot alkalmazzuk, akkor a következő vektor-sorozatot készítjük:

$$z_1 = a_{i+1},\ \ z_2 = P_1 z_1,\ \ldots\ ,\ z_{j+1} = P_j z_j,\ \ q_{i+1} = z_{i+1}.$$

A vetítőmátrixok kiírásával ekkor az

$$r_{j,i+1} = \frac{q_j^T z_j}{q_j^T q_j} \tag{6.8}$$

előállításra jutunk.

A szummás alakot nevezzük a *klasszikus Gram-Schmidt (G-S) ortogonalizációnak,* a szorzatmátrixos változatot pedig *módosított Gram-Schmidt ortogonalizációnak.* Åke Björck a numerikus tulajdonságok vizsgálata során kimutatta, hogy a módosított G-S módszer jobb hibatulajdonságokkal rendelkezik. Újabb eredmények szerint mindkét módszer egyformán jó, ha minden ortogonalizációs lépést kétszer hatjunk végre. Ekkor a kapott normált vektorok ortogonalitása közel gépi pontossággal teljesül.

6.1.1 Feladatok

- Igazoljuk a (6.5) formulát!
- Mutassuk meg, hogy a (6.7) és (6.8) formulával adott $r_{j,i+1}$ megegyezik!
- Gyűjtsük az ortogonális vektorokat a $Q = [q_1 q_2 \ldots q_i]$ mátrixba. Vezessük le, hogy $P_i P_{i-1} \ldots P_1 =$ $= I - Q\left(Q^T Q\right)^{-1} Q^T$.
- Legyen $A \in \mathbb{R}^{m \times n}$, ahol A oszlopai lineárisan függetlenek. Ellenőrízzük, hogy $I - A\left(A^T A\right)^{-1} A^T$ szintén vetítőmátrix és egy vektorra alkalmazva az eredmény olyan vektor lesz, amely A összes oszlopára ortogonális.

6.2. Tétel, QR-felbontás

Legyen $A \in \mathbb{R}^{m \times n}$, ahol A oszlopai lineárisan függetlenek. Ekkor A mindig felírható

$$A = QR \tag{6.9}$$

alakban, ahol Q oszlopai egymásra ortogonális vektorok és R felső háromszög mátrix. Q és R oszlopai az elsővel kezdve rekurzívan felépíthetők.

Bizonyítás. Szükséges $n \le m$, különben nem lehetnének A oszlopai lineárisan függetlenek. Fogjuk fel az A mátrixot úgy, mint ami az $a_1, a_2, \ldots, a_n$ oszlopvektorokból van összeállítva és alkalmazzuk az előző szakaszban megismert G-S ortogonalizációt! Ekkor (6.6) és (6.7) öszevetéséből kapjuk:

$$a_{i+1} = \sum_{j=1}^{i+1} q_j r_{j,i+1},\ \text{ ahol } r_{i+1,i+1} = 1.$$

Az $A = [a_1, a_2, \ldots, a_n],\ Q = [q_1, q_2, \ldots, q_n]$ mátrixokkal ez pedig nem más, mint a (6.9) előállítás, ahol $R = [r_{ij}]$. A G-S ortogonalizációban az $r_{ij},\ i > j$ elemek nem voltak definiálva. Nincs is rájuk szükség, így ezeket az elemeket zérusnak választva (6.9) pontosan teljesül. ■

6.2.1 Feladatok

- A G-S ortogonalizációnál kidolgozható az a változat, amikor a q_j vektorok normáltak, $\|q_j\|_2 = 1$. Írjuk át a formulákat erre az esetre!
- Legyen $D = Q^T Q$, ezzel $\tilde{Q} = QD^{-1/2}$ oszlopvektorai normáltak. (6.9)-ben legyen $A = \tilde{Q}\tilde{R}$. Adjuk meg $\tilde{R}$-et mátrixos alakban és a normált ortogonális vektorok segítségével fejezzük ki $\tilde{r}_{ij}$-t!
- A 3.12 gyakorlatban láttuk, hogy minden x vektor egy σe_1 vektorba vihető tükrözéssel, ahol $|\sigma| = \|x\|_2$. Ha egy ilyen tükrözést alkalmazunk A első oszlopára, akkor a QR-felbontás az első oszlopra megvalósult: $R_1 = I - 2v_1 v_1^T / v_1^T v_1$ ortogonális mátrix, ahol $v = x - \sigma e_1$ és $A = R_1(R_1 A)$. Hogyan folytassuk a tükrözéseket, hogy egy QR-felbontáshoz jussunk?
- Ha rendelkezésünkre áll A egy QR-felbontása, hogyan oldjunk meg egy $Ax = b$ egyenlet-rendszert?

6.3. Példa QR-felbontásra

Elkészítjük az

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 3 \\ 1 & 1 & -2 \\ 2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$$

mátrixra a QR-felbontás nemnormált változatát. Induláskor $q_1 = a_1$ és $q_1^T q_1 = 10$. A következő vektorhoz $q_1^T a_2 = 6$ és

$$q_2 = a_2 - q_1 \frac{q_1^T a_2}{q_1^T q_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{6}{10}\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 2 \\ -1 \\ 2 \\ -1 \end{bmatrix}.$$

$q_2^T q_2 = \frac{10}{25} = \frac{2}{5}$, ezt az eredményt előállíthatjuk a kapott q_2 vektorból, de úgy is számíthatjuk, hogy észrevesszük: $q_2^T q_2 = \left(a_2^T - \frac{q_1^T a_2}{q_1^T q_1} q_1^T \right)\left(a_2 - q_1 \frac{q_1^T a_2}{q_1^T q_1} \right) = a_2^T a_2 - \frac{\left(q_1^T a_2\right)^2}{q_1^T q_1} = 4 - \frac{36}{10} = \frac{2}{5}$. A harmadik vektor előállításához $q_1^T a_3 = 5$ és $q_2^T a_3 = -2$, ezekkel a harmadik vektor és a QR-felbontás:

$$q_3 = a_3 - q_1 \frac{q_1^T a_3}{q_1^T q_1} - q_2 \frac{q_2^T a_3}{q_2^T q_2} = \begin{bmatrix} -1 \\ 3 \\ -2 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} + \frac{2 \cdot 5}{2 \cdot 5}\begin{bmatrix} 2 \\ -1 \\ 2 \\ -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \\ -2 \end{bmatrix},$$

$$A = \begin{pmatrix} 1 & 2/5 & 1/2 \\ 2 & -1/5 & 1 \\ 1 & 2/5 & -1/2 \\ 2 & -1/5 & -1 \end{pmatrix}\begin{pmatrix} 1 & 3/5 & 1/2 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & -2 \end{pmatrix}\begin{pmatrix} 1 & 3/5 & 1/2 \\ 0 & 1/5 & -1 \\ 0 & 0 & 1/2 \end{pmatrix}.$$

6.3.1 Feladat

Készítsük el a következő mátrix QR-felbontását: $\begin{bmatrix} 2 & 6 & 5 \\ -1 & -4 & 1 \\ -1 & -2 & -3 \end{bmatrix}$.

6.4. Az Arnoldi-módszer

Ez a *Krilov-bázis* vektorainak G-S ortogonalizációja. A *Krilov-bázis* vektorai $x, Ax, A^2x, \dots$, ahol $x \neq 0$, egyébként tetszőleges induló vektor. Az Arnoldi-módszernél ennek alapján $q_1 = x / \|x\|_2$, a következő vektor Aq_1 és ezt q_1-re q_1-re ortogonalizálva kapjuk q_2-t. Általában q_{i+1}-et úgy kapjuk, hogy az Aq_i vektort ortogonalizáljuk a meglévő q_j vektorokra és az eredményt normáljuk. Belátható, hogy az így nyert q_j vektorok ugyanazt az alteret feszítik ki, mint a Krilov-bázis vektorai. A módszerrel a következő QR-felbontáshoz jutunk:

$$[q_1 \quad Aq_1 \quad Aq_2 \quad \dots \quad Aq_i] = [q_1 \quad q_2 \quad \dots \quad q_{i+1}]R, \tag{6.10}$$

ahol R felső háromszögmátrix. Ebből a sémából szokás elhagyni bal oldalon az első vektort, q_1-et. Ez azt jelenti, hogy a jobb oldalon R első oszlopát is elhagyjuk. Sőt, hogy R helyén négyzetes mátrix maradjon, az utolsó sorát is elhagyjuk. Jelöljük a maradék mátrixot H-val. Ekkor a q_j vektorokból mátrixot képezve a

$$Q = [q_1 \quad q_2 \quad \dots \quad q_i], \quad AQ = QH + h_{i+1,i}q_{i+1}e_i^T \tag{6.11}$$

összefüggésre jutunk, ahol H un. *felső Hessenberg-féle* mátrix. E mátrixok közeliek a felső háromszögmátrixokhoz, azzal a különbséggel, hogy a főátló alatti elemek sem zérusok. A rendszert jobbról az e_i vektorral szorozva kapunk rekurziót q_{i+1} számítására:

$$Aq_i = \sum_{j=1}^{i} h_{ji}q_j + h_{i+1,i}q_{i+1}, \quad h_{ji} = q_j^T Aq_i, \tag{6.12}$$

a $h_{i+1,i}$ elemet abból a feltételből határozhatjuk meg, hogy q_{i+1} normált. Ha $h_{i+1,i} = 0$, akkor a rekurzió megáll és $i < n$ esetén Q oszlopai A egy invariáns alterét feszítik ki.

6.4.1 Feladat

1. Legyen az x kezdővektor A három sajátvektorának az összege. Hány lépés után áll le az Arnoldi-módszer?

7. Az algebrai sajátértékfeladat

Eszerint keresendő egy (λ, y, x) hármas, amelyre teljesül

$$Ax = \lambda x, \qquad y^T A = \lambda y^T, \tag{7.1}$$

ahol λ az $A \in \mathbb{R}^{n \times n}$ mátrix *sajátértéke*, x a *jobboldali* és y a *baloldali sajátvektor*. A sajátértékek a $|\lambda I - A|$ *karakterisztikus polinom* gyökei és a sajátértékhelyeken $\lambda I - A$ szinguláris. A determináns alakból látható, hogy a mátrix hasonlósági transzformáltjának karaketerisztikus polinomja ugyanaz: $|\lambda I - S^{-1}AS| = |S^{-1}(\lambda I - A)S| = |S^{-1}||\lambda I - A||S| = |\lambda I - A|$, tehát a hasonlósági transzformáció a sajátértékeket helyben hagyja.

A mátrixot általában valósnak tekintjük. De mivel valós mátrixnak is lehetnek komplex sajátértékei és sajátvektorai, emiatt sokszor a komplex esetre is gondolni kell.

7.1. Néhány tulajdonság

Az alábbiakban felidézünk néhány a sajátértékfeladattal kapcsolatos ismeretet.

7.1.1 Legalább 1 sajátpár létezése

Minden λ_i sajátértékhez tartozik legalább egy jobb- és baloldali sajátvektor.

Mert $\lambda_i I - A$ és $\lambda_i I - A^T$ magtere legalább 1-dimenziós (ui. nemcsak a null-vektorból áll). ■

7.1.2 Lineáris függetlenség

Különböző sajátértékekhez tartozó sajátvektorok lineárisan függetlenek.

Ennek a bizonyítása indirekt módon történhet. Feltesszük két különböző sajátértékhez tartozó sajátvektorról, hogy lineárisan összefüggők. Ekkor ellentmondásra jutunk, mert két vektor úgy lehet linerárisan összefüggő, hogy egyirányúak, ekkor viszont nem lehetnek a sajátértékek különbözők. ■

7.1.3 Különböző sajáértékhez tartozó bal és jobb sajátvektorok ortogonalitása

Legyen v_i a λ_i sajátértékhez tartozó bal sajátvektor, u_j pedig a λ_j sajátértékhez tartozó jobb sajátvektor, $i \neq j$. Ekkor $v_i^T u_j = 0$.

Bizonyítás. Tekintsük a következő kifejezést: $v_i^T A u_j = \lambda_i v_i^T u_j = \lambda_j v_i^T u_j$, ahol az egyik esetben a bal, a másik esetben pedig a jobb sajátvektor tulajdonságot alkalmaztuk. Kapjuk, hogy $(\lambda_i - \lambda_j) v_i^T u_j = 0$, s ebből következik az állítás, mert $\lambda_i \neq \lambda_j$. ■

7.1.4 Következmény

Ha minden sajátérték különböző, akkor a sajátvektorokat egy $X = [x_1 x_2 \ldots x_n]$ és $Y = [y_1 y_2 \ldots y_n]$ mátrixba rendezve kapjuk:

$$AX = X\Lambda, \qquad Y^T A = \Lambda Y^T. \tag{7.2}$$

A lineáris függetlenség miatt X és Y invertálhatók, így $X^{-1}AX = \Lambda = Y^T A Y^{-T}$, ahol a transzponált inverzét jelöltük $-T$-vel. Írhatjuk: $Y^T = D^{-1}X^{-1}$, ahol D egy nemszinguláris diagonálmátrix és szerepe csupán annyi, hogy az y_i vektorok hosszát skálázza. Tehát az általánosság megszorítása nélkül vehetjük: $Y^T = X^{-1}$, a sajátvektorok saját-altereket adnak, a vektorok hossza tetszőleges. A kapott alak mutatja, hogy ekkor a mátrix *hasonlósági transzformációval diagonalizálható*.

7.1.5 Schur tétele

Minden négyzetes mátrix unitér hasonlósági transzformációval felső háromszög alakra hozható.

Bizonyítás. Jelölje $R(u) = I - 2uu^H / u^H u$ a Householder tükröző mátrixot. Legyen $x \neq e_1$ egy normált sajátvektor, $\|x\|_2 = 1$, amelyet skálázzunk úgy, hogy az első eleme valós, nempozitív szám legyen. (Ezt mindig elérhetjük, ha a vektort a nemzérus $-\overline{x}_1 / |x_1|$ számmal beszorozzuk.) Ekkor $R(x - e_1)e_1 = x$ és $R(x - e_1)x = e_1$. Feltéve, hogy $Ax = \lambda x$ teljesül,

$$R(x - e_1)AR(x - e_1)e_1 = R(x - e_1)Ax = R(x - e_1)\lambda x = \lambda e_1.$$

Mivel $R(x - e_1)$ involutórius (azaz megegyezik az inverzével), hasonlósági transzformációt végeztünk, ahol az első oszlopvektor az e_1 vektor λ-szorosába ment át, amivel a felső háromszög alak az első sorban és oszlopban előállt. Az eljárást folytatva az eggyel kisebb méretű jobb alsó blokkokra, végül a kívánt alakra jutunk. ■

Megjegyzés. Ha $x = e_1$, akkor A első oszlopa már λe_1 alakú. A felső háromszög-alakra hozás egy lépése egyben *deflációs módszer*, mert olyan 1-gyel kisebb méretű mátrixot kapunk a jobb alsó sarokban, amelynek a sajátértékei a megtalált λ-t kivéve megegyeznek a kiinduló mátrixéval.

A Schur-tétel segítségével további fontos tulajdonságok láthatók be.

7.1.6 Tétel, diagonalizálhatóság unitér hasonlósági transzformációval

Az A mátrix normális $\Leftrightarrow$ A unitér hasonlósági transzformációval diagonalizálható.

Bizonyítás. $A \in \mathbb{C}^{n\times n}$ normális, ha teljesül $AA^H = A^H A$, ^{H} a transzponált konjugáltat jelöli.

$\Leftarrow$: Tfh Legyen $A = U\Lambda U^H$, innen $AA^H = U\Lambda U^H U\overline{\Lambda} U^H = U\Lambda\overline{\Lambda} U^H = U\overline{\Lambda} U^H U\Lambda U^H = A^H A$.

$\Rightarrow$: Ha A normális, akkor bármely unitér hasonlósági transzformáltja is az. Legyen a Schur-tétel alapján $B = U^H AU$ felső háromszögmátrix, ekkor $BB^H = B^H B$. Ennek az 1,1-indexű eleme:

$$e_1^T BB^H e_1 = \|B^H e_1\|_2^2 = \sum_{j=1}^{n} |b_{1j}|^2 = e_1^T B^H Be_1 = \|Be_1\|_2^2 = |b_{11}|^2,$$

azaz B első sorának kettes normája megegyezik az első oszlopéval. Ez csak úgy lehetséges, ha $b_{1j} = 0,\ j = 2,\dots,n$. Az eljárást folytatva az eggyel kisebb méretű jobb alsó blokkal, minden sorra azt kapjuk, hogy csak főátlóbeli elem lehet nemzérus. ■

A tétel következménye, hogy a valós szimmetrikus mátrixok ortogonális, az hermitikus mátrixok pedig unitér hasonlósági transzformációval diagonalizálhatók. E mátrixok sajátértékei mindig valósak.

7.1.7 Tétel

Az egyszeres sajátértékhez tartozó y bal- és x jobboldali sajátvektorok skaláris szorzata nemzérus: $y^H x \neq 0$.

Hozzuk a mátrixot unitér hasonlósági transzformációval felső háromszög alakra: $B = U^H AU$. Ekkor a sajátvektorok átmennek az $y \to U^H y$ és $x \to U^H x$ vektorokba, ahonnan látható, hogy skaláris szorzatuk nem változik meg. Az általánosság megszorítása nélkül feltehetjük, hogy

$$B = \begin{bmatrix} \lambda & b^T \\ 0 & C \end{bmatrix}$$

alakú, ahol λ az x és y vektorhoz tartozó sajátérték. A transzformálás után x átment e_1-be, a bal oldali vektor pedig legyen $y^H U = \begin{bmatrix} \overline{\eta} & y_2^H \end{bmatrix}$. Ha ezzel balról szorozzuk B-t , akkor

$$\begin{bmatrix} \overline{\eta} & y_2^H \end{bmatrix} \begin{bmatrix} \lambda & b^T \\ 0 & C \end{bmatrix} = \begin{bmatrix} \overline{\eta}\lambda & \overline{\eta} b^T + y_2^H C \end{bmatrix} = \lambda \begin{bmatrix} \overline{\eta} & y_2^H \end{bmatrix},$$

ahonnan $\overline{\eta} b^T + y_2^H C = \lambda y_2^H$. Ha itt $\overline{\eta} = y^H x$ zérus volna, akkor a jobb alsó C blokknak λ sajátértéke volna. Ez ellentmondana annak, hogy λ egyszeres sajátérték. ■

7.1.8 Jordan-blokkok

A

$$J(\mu) = \begin{bmatrix} \mu & 1 & & \\ & \mu & \ddots & \\ & & \ddots & 1 \\ & & & \mu \end{bmatrix} \in \mathbb{C}^{k \times k} \tag{7.3}$$

alakú mátrixot *Jordan-blokknak* nevezzük. Ez a hasonlósági transzformációval nem diagonalizálható mátrixok prototípusa. Ránézésre látható, hogy a karakterisztikus polinomja $|\lambda I - J(\mu)| = (\lambda - \mu)^k$, tehát $\lambda = \mu$ k-szoros gyök. A sajátértékhelyen a karakterisztikus mátrix rangvesztése csak 1, mert a bal alsó sarokelemhez tartozó aldetermináns éppen az átló feletti -1-esek szorzata lesz, így a rang csak eggyel csökken. Következésképp 1 jobb és baloldali sajátvektor van, e_1 és e_k^T, amelyek skaláris szorzata 0, ha $k > 1$.

A sajátérték karakterisztikus polinombeli multiplicitását *algebrai multiplicitásnak,* m_A nevezzük. A sajátértékhez tartozó sajátvektorok által kifeszített altér dimenziója pedig a sajátérték *geometriai multiplicitása,* m_G. A fenti Jordan-blokknál $m_A = k$ és $m_G = 1$.

Bizonyítás nélkül megemlítjük, hogy általában a mátrixok hasonlósági transzformációval *Jordan-féle kanonikus alakra* hozhatók, ahol minden sajátértékhez egy vagy több Jordan-blokk tartozik, amelyek a főátló mentén helyezkednek el. Lehetséges 1×1-es Jordan-blokk, az egyszeres sajátértékeknek például ez van. Könnyen belátható, hogy a sajátértékhelyen a karakterisztikus mátrix rangvesztése egyenlő m_G-vel, ami a sajátértékhez tartozó Jordan-blokkok száma, a sajátérték algebrai multiplicitása m_A viszont egyenlő a hozzátartozó Jordan-blokkokban lévő átlóelemek számával.

7.1.9 Feladatok

1. Legyen A olyan felső Hessenberg mátrix, amelynek minden átló alatti eleme nemzérus. Mutassuk meg, hogy ennek a mátrixnak minden sajátértékéhez csak 1 Jordan-blokk tartozhat.

2. Mutassuk meg, ha A sajátértékei a λ_i-k, akkor A^{-1} sajátértékei $1/\lambda_i$-k.

7.2. A sajátértékek lokalizációja

Mivel még valós mátrixoknak is lehetnek komplex sajátértékei, emiatt a komplex síkon kell megadni olyan tartományokat, ahol a sajátértékek lehetnek. Egy ilyen becsléssel már megismerkedtünk a normákkal foglalkozó 2.8 szakaszban, mely szerint a spektrál sugár nem nagyobb, mint a mátrix valamely indukált normája. Így egyik sajátérték sem lehet nagyobb abszolút értékben, mint például $\|A\|_1$ vagy $\|A\|_\infty$. Ennél pontosabb becslést tesz lehetővé

7.2.1 Gersgorin tétele

Legyen az i-edik K_i Gersgorin-kör középpontja a_{ii}, sugara pedig $r_i = \left\| e_i^T (A - a_{ii} I) \right\|_\infty$, ami nem más, mint a mátrixban az i-edik sorelemek abszolút összege a diagonálelem kivételével. A tétel szerint az A mátrix sajátértékei a Gersgorin-körök egyesített halmazában vannak.

Bizonyítás. Tekintsük az $Ax = \lambda x$ egyenlet i-edik sorát, ahol x sajátvektor, λ a hozzátartozó sajátérték, és $|x_i| = \|x\|_\infty$. Kissé átrendezve: $\lambda - a_{ii} = \sum_{j=1, j\neq i}^{n} \frac{a_{ij} x_j}{x_i}$, ahonnan $|\lambda - a_{ii}| \leq \sum_{j=1, j\neq i}^{n} \left| \frac{a_{ij} x_j}{x_i} \right| \leq$

$\leq \sum_{j=1, j\neq i}^{n} |a_{ij}| = r_i$. Minden sajátértékre felírhatunk egy ilyen összefüggést, ez adja az állítást. ■

7.2.2 Második tétel

Ha a Gersgorin köröknek vannak diszjunkt részhalmazai, akkor minden ilyen részhalmazban annyi sajátérték található, amennyi a hozzá tartozó Gersgorin körök száma.

Bizonyítás. Fel kell használnunk azt az itt nem bizonyított eredményt, hogy a mátrix sajátértékei a mátrixelemek folytonos függvényei. Bontsuk a mátrixot két részre, és képezzük az $A(\varepsilon) = D + \varepsilon A_1$ mátrixot, ahol D a főátlót tartalmazó diagonálmátrix, A_1 pedig a nemdiagonális rész. Ha most $\varepsilon = 0$, akkor minden kör sugara zérus. Ha ε 1-hez tart, akkor minden sajátérték kifuthat a középpontból, de a folytonosság miatt nem ugorhat át egy másik – a sajátjától diszjunkt körhalmazba. ■

7.2.3 Példa

A Gersgorin-tétel alkalmazását kombinálhatjuk diagonálmátrixszal készített hasonlósági transzformációval. Ezzel változtatni tudjuk a körök sugarát, és egyszerűen készíthetünk a célnak megfelelő becslést. Például mutassuk meg, hogy a

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

mátrixnak nincs zérus sajátértéke!

Az első Gersgorin kör középpontja 8, sugara szintén 8, így ez a kör tartalmazza a zérust. A többi kör nem. Alkalmazzuk a $D^{-1}AD$ hasonlósági transzformációt, ahol $D = \mathrm{diag}(2 \quad 1 \quad 1)$:

$$D^{-1}AD = \begin{bmatrix} 8 & 5/2 & 3/2 \\ 2 & 4 & 1 \\ 2 & 2 & 5 \end{bmatrix}$$

ezzel a kívánt célt elértük, mert az első kör sugara 4-re csökkent és a másik két kör továbbra sem tartalmazza a zérust. Figyeljük meg, milyen sor és oszlopban lesz változás, ha az alkalmazott diagonálmátrixban csak egy elem különbözik 1-től!

7.2.4 Feladatok

Bizonyítsuk be:

1. A mátrix főátló-domináns, ha a Gersgorin körök nem tartalmazzák a zérust.
2. A főátló-domináns mátrixok invertálhatók, mert nincs zérus sajátértékük.
3. Az i- és j-edik sorok és oszlopok cseréje a diagonál-dominanciát nem változtatja meg.

4. A mátrix rangja legalább akkora, mint azon Gersgorin körök száma, amelyek nem tartalmazzák a zérust.
5. A baloldali sajátvektorok segítségével is készíthetünk Gersgorin-köröket a mátrix oszlopai szerint.
6. Döntsük el Gersgorin tétele és diagonálmátrix hasonlósági transzformáció segítségével, hogy A invertálható-e: $A = \begin{bmatrix} 7 & 6 & -3 \\ 1 & 5 & 1 \\ 4 & -2 & 6 \end{bmatrix}$.

7.3. A karakterisztikus polinom számítása

Tekintsük az un. *Frobenius-féle kisérő mátrixot*:

$$F = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ & 1 & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{bmatrix}. \tag{7.4}$$

Az utolsó oszlopa mentén kifejtve igazolható, hogy $\det(\lambda I - F) = \lambda^n + \sum_{i=0}^{n-1} a_i \lambda^i$. Eszerint a karakterisztikus polinom együtthatóinak előállítása könnyű, ha van valamilyen hasonlósági transzformáció, ami az A mátrixot ilyen alakra hozza. Danyiljevszkij ötelete szerint ez megvalósítható a Gauss-Jordan módszernél megismert egyszerű transzformációs mátrixszal, ha a mátrix első oszlopát nem e_1-be, hanem e_2-be visszük. Legyen tehát az első transzformációs mátrix $T_1 = I + (Ae_1 - e_2)e_2^T$, $A_2 = T_1^{-1} A T_1$ és ekkor a hasonlósági transzformáció eredményeként az első oszlop e_2 lesz:

$$A_2 e_1 = T_1^{-1} A T_1 e_1 = \left(I - \frac{(Ae_1 - e_2)e_2^T}{e_2^T A e_1} \right) A \left(I + (Ae_1 - e_2)e_2^T \right) e_1 = \left(I - \frac{(Ae_1 - e_2)e_2^T}{e_2^T A e_1} \right) A e_1 = e_2 .$$

Általában a k-adik lépésben $T_k = I + (A_k e_k - e_{k+1})e_{k+1}^T$, és a korábbi oszlopvektorok sem romlanak el, mert az előbbihez hasonlóan kapjuk:

$$\left(I - \frac{(A_k e_k - e_{k+1})e_{k+1}^T}{e_{k+1}^T A_k e_k} \right) A_k \left(I + (A_k e_k - e_{k+1})e_{k+1}^T \right) e_l = e_{l+1}, \quad l \le k .$$

Egy transzformációs lépés végrehajthatóságának feltétele, hogy a diagonálelem alatti elem legyen zérustól különböző. Ha nem így volna, sor-cserével mozgassunk egy átló alatti nemzérus elemet az átlóelem alá és hajtsuk végre a hasonló oszlopok cseréjét is a hasonlósági transzformáció megőrzése végett. Az algoritmus $n-1$-edik lépésében a (7.4) alakhoz jutunk.

Ha a mátrix háromátlójú, a karakterisztikus polinom egyszerű rekurzióval számolható. Például a

$$\begin{vmatrix} \lambda - d_1 & -c_1 & & \\ -a_1 & \lambda - d_2 & \ddots & \\ & \ddots & \ddots & -c_{n-1} \\ & & -a_{n-1} & \lambda - d_n \end{vmatrix}$$

determináns rekurziója

$$p_{i+1}(\lambda) = (\lambda - d_{i+1}) p_i(\lambda) - a_i c_i p_{i-1}(\lambda), \; p_0 = 1, \; p_1 = \lambda - d_1, \tag{7.5}$$

ahol $p_i(\lambda)$ a bal felső i-edrendű blokk determinánsa. Az $i+1$-edrendű determinánst az $i+1$-edik oszlop szerinti kifejtéssel kapjuk és az eredmény a kapott rekurzió. A rekurzióval a polinom helyettesítési értékét is könnyen számolhatjuk.

A rekurziót meg lehet csinálni a felső Hessenberg-mátrix karakterisztikus polinomjára is. Legyen például $p_3 = 1$ és alulról felfelé oldjuk meg a következő egyenletrendszert:

$$\begin{bmatrix} \lambda-2 & 1 & 3 \\ 2 & \lambda+1 & 2 \\ 0 & 2 & \lambda-1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} p(\lambda) \\ 0 \\ 0 \end{bmatrix}.$$

Az utolsó sorból $p_2(\lambda) = (1-\lambda)/2$, a második sorból pedig $2p_1(\lambda) + (\lambda+1)p_2(\lambda) + 2 = 0$, ahonnan p_1 kifejezhető. Ezekkel az első sor adja $p(\lambda)$ kifejezését. A mátrix determinánsa akkor lesz zérus, ha $p(\lambda)$ zérus, tehát $p(\lambda)$ gyökei megegyeznek a mátrix sajátértékeivel. Figyeljük meg, $p(\lambda) = 0$ esetén $p_1(\lambda), p_2(\lambda), p_3(\lambda)$ a λ sajátértékhez tartozó sajátvektor elemei.

7.3.1 Feladat

Igazoljuk, hogy a 2×2-es A mátrix sajátértékei: $\lambda_{1,2} = \frac{a_{11}+a_{22}}{2} \pm \sqrt{\left(\frac{a_{11}-a_{22}}{2}\right)^2 + a_{12}a_{21}}$.

7.4. Tétel

Minden mátrix unitér hasonlósági transzformációval felső Hessenberg-alakra hozható.

Bizonyítás. Az első lépésben legyen $u_1 = (A - e_1 e_1^T A)e_1$, $\|u_1\|_2 = |\sigma_1|$, tehát A első oszlopából az átlóelemet elhagyjuk. A hasonlósági transzformációt az $R(u_1 - \sigma_1 e_2)$ tükröző mátrixszal végezzük, ahol a kivonási jegyveszteség elkerülése érdekében σ_1 előjelét aszerint választjuk, hogy u_1/σ_1 második elemének valós része legyen negatív. Ekkor $R(u_1 - \sigma_1 e_2)A$ az első oszlopot $a_{11}e_1 + \sigma_1 e_2$-be viszi, (az első elem változatlan az első oszlopvektor második elemmel kezdődő részét pedig $\sigma_1 e_2$-be tükröztük). Ugyanezzel a tükröző mátrixszal jobbról szorozva az első oszlop már nem fog változni, mert $R(u_1 - \sigma_1 e_2)$ első sora és oszlopa e_1^T és e_1. Ezzel $A_2 = R(u_1 - \sigma_1 e_2)AR(u_1 - \sigma_1 e_2)$ első oszlopa mutatja a Hessenberg-alakot. A következő lépésben az imént látottakat alkalmazzuk A_2 eggyel kisebb méretű jobb alsó blokkjára. Az eljárást folytatva végül a kívánt teljes Hessenberg-alakra jutunk. ■

7.5. Iterációs módszerek

7.5.1 A hatványiteráció

A módszer azon az észrevételen alapul, hogy k növekedésével $A^k x_0$-ben a legnagyobb sajátértékhez tartozó komponens fog felerősödni. A konvergenciára kimondhatjuk a következő tételt:

Tegyük fel A n-edrendű valós vagy komplex mátrix és a sajátértékeire teljesül

$$|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \ldots \ge |\lambda_n|.$$

Továbbá a mátrix egyszerű struktúrájú, azaz annyi sajátvektora van, mint a mátrix rendje. Ekkor a spektrálfelbontás $A = \sum_{i=1}^{n} \lambda_k u_k v_k^T$, ahol v_k, u_k a bal és jobb sajátvektorok és kifejthető a sajátvektorok szerint: $x_0 = \sum_{k=1}^{n} \alpha_k u_k$. Ekkor

$$\lim_{m\to\infty}\frac{1}{\lambda_1^m}A^m x_0=\alpha_1 u_1 \tag{7.6}$$

Bizonyítás. A módszer szerint képezzük az $x_m = Ax_{m-1} = \sum_{k=1}^{n}\alpha_k\lambda_k^m u_k$ vektorokat és innen kapjuk $\frac{A^m x_0}{\lambda_1^m}=\sum_{k=1}^{n}\alpha_k\left(\frac{\lambda_k}{\lambda_1}\right)^m u_k$. Az $m\to\infty$ határátmenettel kapjuk az állítást, mivel a többi sajátvektor szorzója zérushoz tart. ■

Látjuk, a konvergencia gyorsaságát $\alpha_1\neq 0$ esetén lényegében a λ_2/λ_1 hányados szabja meg. Az algoritmus:

$$\begin{aligned}
&m=1,2,\ldots-re:\\
&y_{m+1}=Ax_m,\\
&x_{m+1}=\frac{y_{m+1}}{\|y_{m+1}\|}.
\end{aligned}$$

A normának célszerű például a végtelen normát választani. A sajátérték a

$$\lambda_1^{(m)}=\frac{x_m^T y_{m+1}}{x_m^T x_m}=\frac{x_m^T A x_m}{x_m^T x_m}$$

kifejezéssel becsülhető. A hatványmódszerrel a spektrum (= a mátrix sajátértékeinek összessége) szélein lévő egyszeres sajátértékeket kereshetjük sikerrel. Az *inverz hatányiterációval* azonban kereshetjük a spektrum belsejében elhelyezkedő sajátértékeket is. Ekkor az iteráció egy lépésében az

$$x_{m+1}=(\lambda I-A)^{-1}x_m$$

vektort számítjuk. A $(\lambda I-A)^{-1}$ mátrix sajátértékei $1/(\lambda-\lambda_k)$, $k=1,\ldots,k$, innen látható: ez is hatványiteráció, ami a λ paraméter értékéhez legközelebb eső sajátértékhez és a hozzátartozó sajátvektorhoz fog tartani.

A sajátprobléma megoldására az egyik legjobb módszer a QR-módszer. Ekkor elkészítjük az $A=Q_1R_1$ QR-felbontást és a következő mátrix $A_2=R_1Q_1=Q_1^TAQ_1$, tehát egy ortogonális hasonlósági transzformáció eredménye. A k-edik lépésben $A_k=R_{k-1}Q_{k-1}$. Megmutatható, hogy amennyiben a mátrix egyszerű struktúrájú és a sajátvektorok mátrixának van LU-felbontása, akkor a QR-módszer egy felső háromszögmátrixhoz konvergál. A konvergencia még gyorsítható, ha a felbontásokat kombináljuk egy κI mátrixszal való eltolással is, ahol κ a sajátérték egy becslése. Ekkor a QR-módszer konvergencia-sebessége másodrendű, szimmetrikus mátrixoknál harmadrendű lesz.

7.6. A sajátértékfeladattal kapcsolatos egyenlőtlenségek

A Gersgorin-körök ismertetése során már megismerkedtünk ilyen összefüggésekkel. Itt folytatjuk a vizsgálatainkat. Arra vagyunk kiváncsiak, hogyha van egy közelítő (λ,u) sajátpárunk, mit mondhatunk a jóságának jellemzésére. Egy másik feladat, hogyha a mátrixelemeket kissé megváltoztatjuk (– perturbáljuk), hogyan változik meg a sajátpár?

A következő *jelöléseket* alkalmazzuk: $M=\max_{(i)}|\lambda_i(A)|$, $m=\min_{(i)}|\lambda_i(A)|$ és feltesszük, hogy a mátrix invertálható. Mindig indukált mátrixnormát használunk.

A spektrálsugár és az indukált normák összefüggéséből már ismerjük: $M\le\|A\|$, $1/m\le\|A^{-1}\|$, a kettő összeszorzásából:

$$\frac{M}{m} \le \mathrm{cond}(A) = \|A\| \|A^{-1}\| . \tag{7.7}$$

Ez jelzi számunkra, hogyha abszolút értékben a legnagyobb és legkisebb sajátérték hányadosa nagy, akkor a mátrix kondiciószáma nagy.

7.6.1 Lemma

Legyen $D = \mathrm{diag}(d_1, \ldots, d_n)$ diagonálmátrix, ekkor $\|D\|_p = \max\limits_{(i)} |d_i|, \quad 1 \le p \le \infty$.

Bizonyítás. Legyen $|d_k| \ge |d_i|$ minden i -re. Az indukált norma definiciót alkalmazva

$$\|D\|_p^p = \sup_{(x \neq 0)} \frac{\sum_{i=1}^n |d_i x_i|^p}{\sum_{i=1}^n |x_i|^p} = |d_k|^p \sup_{(x \neq 0)} \frac{\sum_{i=1}^n |x_i d_i / d_k|^p}{\sum_{i=1}^n |x_i|^p} = |d_k|^p ,$$

mert a nevező nagyobb a számlálónál, ha van olyan nemzérus x_i elem, amelyre $|d_i / d_k| < 1$. ■

7.6.2 Tétel, sajátpár jósága

Legyen A egyszerű szerkezetű: $AU = U\Lambda$, ahol U a sajátvektorok mátrixa és Λ a sajátvektorokat tartalmazó diagonálmátrix, továbbá (λ, x) a sajátpár egy közelítése. Ekkor az $r = Ax - \lambda x$ jelöléssel

$$\min_{(i)} |\lambda_i - \lambda| \le \frac{\|r\|}{\|x\|} \mathrm{cond}(U), \quad \| \, . \, \| \ p\text{-norma} . \tag{7.8}$$

Bizonyítás. Ha $\lambda_i = \lambda$ valamely i -re, akkor az állítás igaz. Tegyük fel, $\lambda_i \neq \lambda$, ezzel $A - \lambda I$ invertálható: $x = \left(A - \lambda I\right)^{-1} r = U \left(\Lambda - \lambda I\right)^{-1} U^{-1} r$. A normákra áttérve és az előző lemmát alkalmazva

$$\|x\| \le \frac{\mathrm{cond}(U)}{\min\limits_i |\lambda_i - \lambda|} \|r\| .$$

A kapott egyenlőtlenséget rendezve kapjuk a állítást. ■

7.6.3 Következmény

Hermitikus mátrixokra U unitér, emiatt $\mathrm{cond}_2(U) = 1$ és $\min\limits_{(i)} |\lambda_i - \lambda| \le \|r\|_2 / \|x\|_2$, ami nagyon egyszerűen számolható.

7.6.4 Tétel, alsó becslés cond(U)-ra

Ha A egyszerű szerkezetű és invertálható,

$$\frac{\|A\|}{M} \le \mathrm{cond}(U), \quad \|A^{-1}\| m \le \mathrm{cond}(U), \quad \sqrt{\frac{\mathrm{cond}(A)}{\mathrm{cond}(\Lambda)}} \le \mathrm{cond}(U) \tag{7.9}$$

Bizonyítás. A harmadik összefüggés az első kettő összeszorzásával adódik. Az első egyenlőtlenség az $A = U \Lambda U^{-1}$ normáját képezve adódik, a második pedig az $A^{-1} = U \Lambda^{-1} U^{-1}$ kifejezésből. ■

7.6.5 Feladatok

Mutassuk meg, hogy

1. $\|U\| \le \|AU\| / m$.

2. $\|U^{-1}\| \le \|U^{-1}A^{-1}\| M$.

3. $\mathrm{cond}(U) \le \mathrm{cond}(AU)\mathrm{cond}(\Lambda)$.

7.6.6 Tétel, (Bauer, Fike)

Legyen A egyszerű szerkezetű és E egy ugyanolyan méretű mátrix. Ha μ az $A+E \in \mathbb{C}^{n\times n}$ egy sajátértéke és $AU = U\Lambda$, akkor

$$\min_{(i)} |\lambda_i - \mu| \le \|E\|_p \,\mathrm{cond}_p(U) . \tag{7.10}$$

Bizonyítás. Tegyük fel $\mu \notin \{\lambda_i(A)\}$, mert különben igaz az állítás. Mivel μ sajátérték, következik, hogy $U^{-1}(A+E-\mu I)U = \Lambda - \mu I + U^{-1}EU$ szinguláris, ahonnan átrendezéssel $I + (\Lambda - \mu I)^{-1} U^{-1} EU$ adódik. Ez az utóbbi mátrix pedig csak akkor lehet szinguláris, ha $(\Lambda - \mu I)^{-1} U^{-1} EU$ -nek van egy 1 abszolút értékű sajátértéke, amiből $1 \le \left\|(\Lambda - \mu I)^{-1} U^{-1} EU\right\|_p \le \left\|(\Lambda - \mu I)^{-1}\right\|_p \|E\|_p \,\mathrm{cond}_p(U)$. Ezt rendezve kapjuk az állítást. ■

7.6.7 Tétel, inverz perturbáció

Legyen (λ, x) egy közelítő sajátpár, $r = Ax - \lambda x$. Ekkor az $E = -\dfrac{rx^H}{\|x\|_2^2}$, $\|E\|_2 = \dfrac{\|r\|_p}{\|x\|_p}$, $p = 2, F$ (F a Frobenius-norma) mátrixszal a (λ, x) sajátpár az

$$(A+E)x = \lambda x \tag{7.11}$$

egyenlet pontos megoldása.

Bizonyítás. $\left(A - \dfrac{rx^H}{x^H x}\right)x = Ax - r = \lambda x$. ■

Például, ha $\|r\|_2 / \|x\|_2 \approx 10^{-9}$, a mátrix elemei 1 körüliek, akkor (λ, x) pontos megoldása egy mátrixnak, ami A -tól csak a kilencedik jegyében különbözik. Ha A -t csak 7 jegyre ismerjük, akkor nincs érteleme tovább folytatni az iterációt.

7.6.8 Tétel, egyszeres sajátérték perturbációja

Legyen (λ, x, y) egy A -hoz tartozó sajáthármas, ahol λ egyszeres sajátérték. Az $A+E$ mátrix sajátértékének megváltozása első rendben

$$\tilde{\lambda} = \lambda + \frac{y^T E x}{y^T x} + \mathcal{O}(\|E\|_2^2) \tag{7.12}$$

és

$$\left|\tilde{\lambda} - \lambda\right| \le \frac{\|y\|_2 \|x\|_2}{\left|y^T x\right|} \|E\|_2 + \mathcal{O}(\|E\|_2^2) . \tag{7.13}$$

Bizonyítás. A második összefüggés az első következménye, ha normákra térünk át. Az első bizonyításához legyen a sajátérték megváltozása μ, a sajátvektoré pedig h:

$$(A+E)(x+h) = (\lambda + \mu)(x+h) .$$

Feltesszük, hogy $E \to 0$ esetén $\mu \to 0$ és $h \to 0$. Beszorzás után a másodrendű tagokat hagyjuk el:

$$Ah + Ex \approx \lambda h + \mu x .$$

Szorozzuk ezt a kifejezést balról az y^T sajátvektorral, ekkor mindkét oldalon az első vektorok is kiesnek és az első bizonyítandó összefüggéshez adódik

$$\mu \approx \frac{y^T Ex}{y^T x} . \tag{7.14}$$

Itt a nevező nem lehet zérus a 7.1.7 Tétel miatt. (7.13)-ban $\|E\|_2$ szorzója nem más, mint az x és y vektoroknál a bezárt szög koszinuszának reciproka: $\sec\angle(x,y) = \|x\|_2 \|y\|_2 / |y^T x|$. Szokás ezt az értéket a λ *sajátérték kondiciószámának* nevezni. ■

8. A legkisebb négyzetek módszere

8.1. Egy illesztési feladat.

Gyakran találkozhatunk a következő feladattal: adottak a (t_i, y_i), $i = 0,1,\dots,n$ pontok, ahol a t_i helyhez tartozó y_i értéket valamely merésből kapjuk. A mért függvényértékeket hiba terheli. Az előálló pontsorozatot – vagy annak egy részét - szeretnénk

egy $f(t) = \sum_{j=0}^{n} c_j \varphi_j(t)$ függvénnyel közelíteni (pl. $\varphi_j(t) = t^j$):

$$\sum_{j=0}^{n} c_j \varphi_j(t_i) \approx y_i . \tag{8.1}$$

Ésszerűnek látszik ezt a feladatot úgy megoldani, hogy az eltérések négyzetösszege minimális legyen:

$$\sum_{i=0}^{m} (y_i - \sum_{j=0}^{n} c_j t_i^j)^2 = \min \tag{8.2}$$

vagyis, (6.1)-ben a c_j lineárkombinációs együtthatókat e feltétel szerint keressük. A (6.1) feladat a c_j ismeretlenekre egy lineáris egyenletrendszer, így tekintsük általánosan az

$$Ax = b,\ A \in \mathbb{R}^{m,n},\ x \in \mathbb{R}^n,\ b \in \mathbb{R}^m \tag{8.3}$$

egyenletrendszert, ahol most az együtthatómátrix nem kvadratikus, hanem téglalap alakú, és az sem biztos, hogy mindig van megoldása. A legkisebb négyzetes tulajdonságnak eleget tevő megoldáshoz ($\|b - Ax\|_2^2 = \min$) vizsgáljuk meg a projektor (vetítő) mátrixokat!

8.2. Vetítő mátrixok, projektorok

8.2.1 Definició

A P mátrixot projektor vagy vetítő mátrixnak nevezzük, ha eleget tesz a $P^2 = P$ összefüggésnek.

Innen rögtön következik: $P(I - P) = (I - P)P = 0$. Ha P invertálható, akkor P^{-1}-et a definiáló egyenletre alkalmazva kapjuk: $P = I$. Ugyancsak egyszerű ellenőrízni, hogyha P projektor, akkor $I - P$ is az.

Figyeljük meg: P a transzformált vektort egyszeri alkalmazás után helybenhagyja: $P(Px) = Px$. Ezután akárhányszor alkalmazzuk a projektort, az eredmény ugyanaz marad, ugyanúgy, mint vetítéskor. Mivel P akárhányadik hatványa önmaga, emiatt használatos az *idempotens mátrix* elnevezés.

Példa projektorra: Legyen $A \in \mathbb{R}^{m,n}$, $B \in \mathbb{R}^{m,n}$, $n < m$, ^{T} jelölje a transzponáltat és legyen $B^T A$ invertálható. Ekkor

$$P = P(A, B^T) = A(B^T A)^{-1} B^T$$

olyan projektor lesz, amely A oszlopvektorainak alterébe vetít:

$$P(A, B^T):\ \mathbb{R}^m \to \mathrm{Im}(A).$$

$(I - P)P = 0$ -ból pedig következik: bármely z vektorra $\{I - P^{\mathrm{T}}(A, B^T)\}z \perp \mathrm{Im}(A)$.

8.2.2 Tétel.

Legyen $\mathcal{P}(\mathcal{A})$ az $\mathcal{A} \subset \mathbb{R}^m$ altérbe vetítő projektorok halmaza. Ekkor bármely $x \notin \mathcal{A}$, $Px \neq 0$ vektorra

$$\max_{P \in \mathcal{P}(\mathcal{A})} \frac{x^{\mathrm{T}} Px}{\|Px\|_2} = \|P_s x\|_2,$$

ahol P_s az altérbe vetítő szimmetrikus projektor.

Bizonyítás. Szemléletesen szólva: az x vektorral $P_s x$ zárja be a legkisebb szöget. Ha $P_s, P \in \mathcal{P}(\mathcal{A})$, akkor

$$P_s P = P$$

hiszen P oszlopvektorai az $\mathcal{A}$ altérbe esnek, és ezeket egy másik olyan projektor mindig helybenhagyja, amely ugyanabba az altérbe vetít. Felhasználva a Cauchy-egyenlőtlenséget, adódik

$$\frac{x^{\mathrm{T}} Px}{\|Px\|_2} = \frac{x^{\mathrm{T}} P_s Px}{\|Px\|_2} \leq \frac{\|P_s x\|_2 \|Px\|_2}{\|Px\|_2} = \|P_s x\|_2, \tag{8.4}$$

ahol a maximum még a $\tilde{P}x = \lambda P_s x$, $\lambda > 0$ feltételt kielégítő $\tilde{P}$ projektorok mellett is előáll. ■

8.2.3 Tétel

Ugyanazon altérbe vetítő projektorok között a szimmetrikus projektor egyértelmű.

Bizonyítás. Indirekt. Tegyük fel, P_1 és P_2 két ugyanabba az altérbe vetítő különböző szimmetrikus projektor, ekkor

$$P_1 P_2 = P_2 \quad \Rightarrow \quad P_2 = P_2^{\mathrm{T}} = P_2^{\mathrm{T}} P_1^{\mathrm{T}} = P_2 P_1 = P_1,$$

ahonnan ellentmondásra jutottunk. ■

8.2.4 Tétel

Az x vektor $\mathcal{A}$ altértől való távolsága kettes normában:

$$\|(I - P_s)x\|_2, \quad P_s \in \mathcal{P}(\mathcal{A}).$$

Bizonyítás. Minthogy $P_s x$ zárja be a legkisebb szöget x -szel, így a $P_s x$ irány mentén található az a pont $\mathcal{A}$ -ban, amely legközelebb van x végpontjához. Keressük tehát azt a λ -t, amelyre $x - \lambda P_s x$ norma-négyzete

$$\|x - \lambda P_s x\|_2^2 = x^{\mathrm{T}} x - 2\lambda x^{\mathrm{T}} P_s x + \lambda^2 x^{\mathrm{T}} P_s x$$

minimális. Deriválással kapjuk, hogy a minimum helye $\lambda = 1$ -nél van. Tehát $(I - P_s)x \perp \mathcal{A}$ és a kettes normája az x vektor $\mathcal{A}$ altértől való távolsága. ■

8.3. Mátrixok általánosított inverze, a pszeudoinverz

Mátrixok általánosított inverzét akkor értelmezzük, ha az inverzük nem létezik. A pszeudoinverz a lineáris egyenletrendszernek azt a megoldását adja, amelyre az eltérés vektor, más szóval reziduum,

$r = b - Ax$ kettes normája minimális. Ha több megoldás is van, akkor a legkisebb kettes normájú megoldást szolgáltatja. Emlékeztetőül két kis lemma felidézésével kezdjük.

8.3.1 Lemma.

Legyenek az L mátrix oszlopai lineárisan függetlenek. Akkor $LB = LC$ -ből $B = C$ következik.

Bizonyítás. Átrendezve $L(B - C) = 0$. L oszlopainak bármely lineáris kombinációja a lineáris függetlenség miatt csak akkor lesz zérusvektor, ha $B - C$ oszlopvektorai zérusok. ■

8.3.2 Lemma.

Legyen $A \in \mathbb{R}^{m,n}$. Ekkor $A^T A$ pozitív szemidefinit. Ha A oszlopai lineárisan függetlenek, azaz A oszloprangú, akkor $A^T A$ pozitív definit.

Bizonyítás. Legyen $y = Ax$, ekkor $x^T A^T Ax = y^T y \geq 0$. Ha A oszloprangú, az előző lemma alapján $y = 0$-ból következik $x = 0$, így ez esetben $A^T A$ pozitív definit. ■

8.3.3 A pszeudoinverz

A továbbiakban megmutatjuk, hogy minden mátrixhoz létezik pszeudoniverz, vagy Moore-Penrose féle általánosított inverz A^+, mely a közönséges inverzzel egyenlő, ha a mátrix nemszinguláris, egyéb esetben viszont minimális kettes norma tulajdonságokkal rendelkezik. Ezt a mátrix inverzet a következő tulajdoságok definiálják:

$$1.\ AA^+A = A, \qquad 2.\ A^+AA^+ = A^+,$$

$$3.\ AA^+ \text{ hermitikus}, \quad 4.\ A^+A \text{ hermitikus}.$$

A definició komplex mátrixokra vonatkozik, mi most valós esetben a hermitikus tulajdonság helyett a szimmetrikusságot követeljük meg. Vegyük észre: Ha az első egyenletet jobbról, vagy balról szorozzuk A^+-tel, az adódik, hogy AA^+ és A^+A projektorok és a 3. és 4. feltétel szerint még szimmetrikusak is. Az első feltétel alapján AA^+ $\mathrm{Im}(A)$ -ba vetít, az $A(I - A^+A) = 0$ alakból pedig azt látjuk, hogy $I - A^+A$ a $\ker(A)$ -ba vetítő szimmetrikus projektor.

Tegyük fel: $A = LU$, ahol L és U közbülső mérete $r = \mathrm{rang}(A)$, $L \in \mathbb{R}^{m,r}$, $U \in \mathbb{R}^{r,n}$. A továbbiakban az LU rang-faktorizáció ismeretében előállítjuk a pszeudoinverzet.

8.3.4 Tétel, a pszeudoinverz előállítása

Legyen $A = LU$ egy rang faktorizáció. Akkor egyértelműen létezik $A^+ = U^+L^+$, ahol $L^+ = (L^T L)^{-1} L^T$ és $U^+ = U^T (UU^T)^{-1}$.

Bizonyítás. Az egyértelműséget indirekt úton bizonyítjuk. Tegyük fel van kettő: A_1^+ és A_2^+. Ekkor a szimmetrikus projektor egyértelműsége miatt $A_1^+A = A_2^+A$ illetve $AA_1^+ = AA_2^+$. Ezeket, és a definiáló egyenleteket felhasználva adódik

$$A_1^+ = A_1^+AA_1^+ = A_2^+AA_1^+ = A_2^+AA_2^+ = A_2^+,$$

amivel ellentmondásra jutottunk. A továbbiakban megkonstruáljuk a pszeudoinverzet.

Vegyük észre: $\mathrm{Im}(A) = \mathrm{Im}(L)$, így ebbe az altérbe vetítő egyértelmű szimmetrikus projektor $AA^+ = LL^+ = L(L^T L)^{-1} L^T$, és $L^+ = (L^T L)^{-1} L^T$ következik a 8.3.1 Lemmából. Hasonlóan az $\mathrm{Im}(A^T) = \mathrm{Im}(U^T)$ altérbe vetítő egyértelmű szimmetrikus projektor $A^+A = A^T (A^+)^T =$

$= U^T (U^+)^T = U^+U = U^T (UU^T)^{-1} U$, ahonnan $U^+ = U^T (UU^T)^{-1}$. Ezek alapján $L^+ L = UU^+ = E_r$ r -edrendű egységmátrixok, és ezzel

$$AA^+ = LL^+ = LUU^+L^+ \text{ és } A^+A = U^+U = U^+L^+LU,$$

Ebből kiolvasható, hogy $A^+ = U^+L^+$. ■

8.3.5 Megjegyzések

Ha A oszloprangú, akkor $L = A$ és $U = I_n$ megfelelő választás, és $A^+ = (A^T A)^{-1} A^T$ következik. Az $A = QR$ faktorizációnál kapjuk: $A^+ = R^{-1} Q^T$. Ha A sorrangú, akkor $L = I_m$ és $U = A$ a megfelelő választás, amivel $A^+ = A^T (AA^T)^{-1}$. Ha most $A^T = QR$, akkor $A^+ = Q(R^T)^{-1}$. Végül, ha A rangja kisebb, mint a legkisebb mérete, azaz, vannak lineárisan összefüggő sorok és oszlopok, akkor mindkét oldal felől végzett ortogonalizációval elérhető az $A = Q_1 B Q_2$ alak, ahol Q_1, Q_2 ortogonális mátrixok és B felső bidiagonális mátrix. Ekkor $A^+ = Q_2^T (B)^{-1} Q_1^T$. Mátrixoknál a rang numerikus meghatározása néha nagyon kényes feladat.

8.3.6 Tétel, lineáris egyenletrendszer megoldhatósága

Legyen P $\mathrm{Im}(A)$ -ba vetítő projektor. Ekkor az $Ax = b$ egyenletrendszer akkor és csak akkor konzisztens (megoldható), ha $Pb = b$.

Bizonyítás. Szükségesség. Ha a rendszer megoldható, akkor $b \in \mathrm{Im}(A)$ és $Pb = b$ -nek teljesülni kell. Az elégségességhez válasszuk az AA^+ szintén $\mathrm{Im}(A)$ -ba vetítő projektort, amelyre $AA^+b = b$ teljesül. Innen kiolvasható, hogy $x = A^+b$ egy megoldás. ■

8.3.7 Tétel, a pszeudoinverzes megoldás tulajdonságai

Az $Ax = b$ lineáris egyenletrendszer általános megoldása a pszeudoinverz segítségével a következőképp állítható elő:

$$x = x_p + x_h = A^+b + (I - A^+A)t, \quad t \in \mathbb{R}^n, \tag{8.5}$$

ahol x_p egy partikuláris megoldás és x_h a homogén egyenlet általános megoldása. Ha a rendszer megoldható, akkor A^+b egy partikuláris megoldás és $(I - A^+A)t$ a homogén egyenlet általános megoldása. Ha a rendszer inkonzisztens, akkor A^+b az a legkisebb négyzetes megoldás, melyre $b - AA^+b$ kettes normája minimális. Minden esetben A^+b a minimális kettes normájú megoldás.

Bizonyítás. A 8.1.4 Tétel alapján $\left\| b - AA^+b \right\|_2$ a b vektor $\mathrm{Im}(A)$ -tól való távolsága. Továbbá vegyük észre, (8.5)-ben két ortogonális vektor van, mert A^+A az első vektort a pszeudoinverz tulajdonságok miatt helyben hagyja, a másodikat pedig zérusba viszi. Emiatt írhatjuk: $\|x\|_2^2 = \left\|A^+b\right\|_2^2 + \left\|(I - A^+A)t\right\|_2^2$, ami akkor a legkisebb, ha $t = 0$, vagy $I - A^+A = 0$. ■

8.4. Feladatok

1. Legyen $A = LU$ egy rang-faktorizáció. Ekkor írjuk fel azt a szimmetrikus vetítőmátrixot, amely $\mathrm{Im}(A)$ -ba vetít.
2. Írjuk fel az A mátrix null-terébe vetítő szimmetrikus projektort! Adjuk meg az x vektor $\mathrm{Nul}(A)$ -tól való távolságát!

3. Egy egyenes áthalad az r_0 és r_1 ponton. Adjuk meg az x vektor és ez az egyenes távolságát!
4. Igazoljuk, hogyha a mátrix invertálható, akkor a pszeudoinverze megegyezik az inverzével.
5. $A^T = \begin{bmatrix} 2 & -4 & 6 \\ 0 & 5 & -5 \end{bmatrix}$. Írjuk fel az $\mathrm{Im}(A)$ -ba vetítő szimmetrikus projektort!
6. Két sík normálvektorát az 5. feladat sorvektorai adják. Melyik az a szimmetrikus projektor, amely a két sík közös részébe vetít?
7. $r^T = [1 \quad -1 \quad 1]$. r végpontja milyen távol van az előbbi két sík közös részétől?
8. $A = \begin{bmatrix} 2 & 0 & 3 \\ -4 & 5 & -2 \\ 6 & -5 & 5 \end{bmatrix}$, $\mathrm{rang}(A) = 2$. $A^+ = ?$
9. Az előbbi mátrixszal mi lesz $Ax = b$ pszeudoinverzes megoldása, ha $b^T = [1 \quad -1 \quad 1]$?
10. Igazoljuk, hogy $I - A^+ A = 0$, ha A oszlopai lineárisan függetlenek.
11. A pszeudoinverz 4 tulajdonságából vezessük le: $(A^+)^T = (A^T)^+$.
12. Az A mátrix közelítő sajátvektora x. A hozzátartozó λ sajátértéket úgy szeretnénk közelíteni, hogy $\|Ax - \lambda x\|_2$ minimális legyen. Mi lesz ekkor λ kifejezése?

9. Ortogonális polinomok

Gyakran polinommal kell végezni a legkisebb négyzetes illesztést. Ilyenkor speciális módszert készíthetünk ortogonális polinomok segítségével. Később a numerikus integrálási módszereknél is szükségünk lesz az ortogonális polinomokra, így most röviden megismerkedünk velük.

9.1. Függvények skaláris szorzata.

Az f és g függvény skaláris szorzatát a következő utasítással definiáljuk:

$$\left(f,g\right)=\int_a^b f(x)g(x)\alpha(x)dx, \quad \alpha(x)>0 \tag{9.1}$$

ahol $\alpha(x)$-et súlyfüggvénynek nevezzük és feltesszük, hogy a kijelölt integrál létezik. Mi most a polinomok használatával összefüggésben egyszerűbb skalárszorzatot fogunk használni, nevezetesen:

$$\left(f,g\right)=\sum_{i=1}^{m} f(x_i)g(x_i)w_i \tag{9.2}$$

ahol x_i, $i=0,1,\ldots,m$ az illesztés alappontjait, w_i pedig a hozzátartozó súlyokat jelenti. Gyakran $w_i=1$ minden i-re.

Ellenőrízzük, hogy a fenti definiciók rendelkeznek a skaláris szorzat tulajdonságaival!

Természetesen most is igaz, hogy a skaláris szorzat normát definiál:

$$\|f\|^2=\left(f,f\right). \tag{9.3}$$

Ezek után semmi akadálya sincs annak, hogy az x^i, $i=0,1,\ldots$ lineárisan független rendszerből Gram-Schmidt ortogonalizációval ortogonális rendszert készítsünk: így kapjuk az *ortogonális polinomokat*.

9.1.1 Definició.

1-*főegyütthatós* az a polinom, amelynél 1 a legmagasabb fokú tag együtthatója.

9.1.2 Tétel

Legyenek $p_i(x)$, $i=0,1,\ldots$ 1-főegyütthatós i-edrendű polinomok. Ekkor bármely $q(x)$ polinom egyértelműen előállítható a p_i polinomok lineáris kombinációjaként:

$$q(x)=\sum_{j=0}^{n} b_j p_j(x). \tag{9.4}$$

Bizonyítás. Legyen $p_i(x)=x^i+p_{i,i-1}x^{i-1}+\ldots+p_{i,0}$, ekkor a b_j együtthatókat meghatározó lineáris egyenletrendszer:

$$\begin{bmatrix} 1 & p_{10} & p_{20} & \cdots & p_{n0} \\ & 1 & p_{21} & \cdots & p_{n1} \\ & & 1 & \ldots & \vdots \\ & \bigcirc & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_n \end{bmatrix} \tag{9.5}$$

.

Ez alulról felfelé haladva egyértelműen megoldható.

9.1.3 Következmény

Legyenek p_i-k ortogonális polinomok. Akkor p_{n+1} ortogonális minden legfeljebb n-edfokú polinomra.

9.2. Az ortogonális polinomok rekurziója

Az 1-főegyütthatós ortogonális polinomok a $p_0(x)$ és $p_1(x)$ polinomok ismeretében rekurzíve felépíthetők:

$$p_{n+1} = (x - \alpha_{n+1}) p_n - \beta_n p_{n-1}. \tag{9.6}$$

Bizonyítás. Vizsgáljuk az

$$(xp_k, p_n) = (p_k, xp_n)$$

skaláris szorzatot! Az eredmény zérus, ha

$$k + 1 < n \text{ és } n + 1 < k$$

a *9.1.3 Következmény* miatt. Nemzérus az eredmény, ha

$$n - 1 \le k \le n + 1,$$

így xp_n kifejthető a p_{n-1}, p_n, p_{n+1} polinomokkal:

$$xp_n = p_{n+1} + \alpha_{n+1} p_n + \beta_n p_{n-1},$$

ahonnan p_{n+1}-re rendezve kapjuk (9.6)-öt. ■

9.2.1 Tétel

Az α_{n+1} és β_n kifejtési együtthatókra érvényes

$$\alpha_{n+1} = \frac{(xp_n, p_n)}{(p_n, p_n)}, \tag{9.7}$$

$$\beta_n = \frac{(xp_n, p_{n-1})}{(p_{n-1}, p_{n-1})} = \frac{(p_n, p_n)}{(p_{n-1}, p_{n-1})}. \tag{9.8}$$

Bizonyítás. Helyettesítsük xp_n értékét a rekurzióból. Az ortogonalitás miatt α_{n+1} értéke rögtön adódik, ha a (9.6) rekurzió mindkét oldalán skaláris szorzatot képezünk p_n-nel. β_n értéke hasonlóan készül, csak most a skaláris szorzatot p_{n-1}-gyel vesszük. Az átalakításban x-et átvisszük p_{n-1}-hez, és az $xp_{n-1} = p_n + \alpha_n p_{n-1} + \beta_{n-1} p_{n-2}$ 1-gyel kisebb indexű rekurziós összefüggést helyettesítjük:

$$\beta_n = \frac{(xp_n, p_{n-1})}{(p_{n-1}, p_{n-1})} = \frac{(p_n, xp_{n-1})}{(p_{n-1}, p_{n-1})} = \frac{(p_n, p_n)}{(p_{n-1}, p_{n-1})}.$$ ■

9.3. Legkisebb négyzetes közelítés ortogonális polinomokkal

A (9.2) skaláris szorzat mellett az induló polinomok, ha $w_i = 1$ minden i-re

$$p_0 \equiv 1, \quad \|p_0\|^2 = \sum_{j=0}^{m} 1 = m+1. \tag{9.9}$$

A következő polinomot p_1-et keressük $x-\alpha_1$ alakban! Ekkor az ortogonalitás miatt

$$(p_0, p_1) = 0 = (p_0, x-\alpha_1) \;\rightarrow\; (p_0, x) = \alpha_1 (p_0, p_0,)$$

ahonnan

$$\alpha_1 = \frac{1}{m+1} \sum_{j=0}^{m} x_j \,, \tag{9.10}$$

így β_0-t zérusnak vehetjük.

A rekurzióból felépített ortogonális polinomokkal a mért $y_i,\; i=0,1,\ldots,m$ függvényértékek a következőképp közelíthetők:

$$y \approx \sum_{j=0}^{k} p_j(x) \frac{(p_j, y)}{(p_j, p_j)}, \quad (p_j, y) = \sum_{i=0}^{m} p_j(x_i) y_i \,. \tag{9.11}$$

Ez az előállítás formálisan ugyanúgy néz ki, mint a lineáris algebrában egy $\{q_j\}$ ortogonális vektorrendszer szerinti kifejtés: legyen y $m+1$-dimenziós vektor, ekkor

$$y = \sum_{j=1}^{k} \frac{q_j q_j^T y}{q_j^T q_j} . \tag{9.12}$$

A (9.11)-ben látható $P_k = \sum_{j=0}^{k} p_j(x) p_j(t) / (p_j, p_j)$ kifejezés is szimmetrikus projektor, így a (9.11) közelítés rendelkezik a legkisebb négyzetes tulajdonsággal: $\|(I-P_k)y\|$ a $p_j,\; j=0,\ldots,k$ polinomok által kifeszített altértől való távolságot jelenti.

9.3.1 Példa

Ortogonális polinomokkal állítsuk elő azt az elsőfokú polinomot, amely az alábbi pontsort legkisebb négyzetesen közelíti:

x_i	-1	0	1	2
y_i	1	2	2	4

Megoldás. Először előállítjuk az ortogonális polinomokat. $p_0(x) = 1$, innen $(p_0, p_0) = 4$. Következik $p_1(x) = x - \alpha_1$, ahol $\alpha_1 = (xp_0, p_0)/(p_0, p_0) = \sum_{j=0}^{3} x_j / (p_0, p_0) = 1/2$. Még meg kell állapítanunk $p_1(x)$ önmagával vett skaláris szorzatát: $(p_1, p_1) = \sum_{j=0}^{3} (x_j - \alpha_1)^2 = \; = \frac{1}{4}(9+1+1+9) = 5$. Ezzel a legkisebb négyzetesen közelítő elsőfokú polinom:

$$P_1(x) = \frac{(p_0, y)}{(p_0, p_0)} p_0 + \frac{(p_1, y)}{(p_1, p_1)} p_1 = \frac{9}{4} + \frac{1}{5} \cdot \frac{1}{2}(-3-2+2+12)(x-\frac{1}{2}) = \frac{9}{4} + \frac{9}{10}(x-\frac{1}{2}).$$

9.4. Feladatok

1. Bizonyítsuk be, hogyha az alappontok $x=0$-ra szimmetrikusan helyezkednek el, akkor $\alpha_j=0,\ i=1,2,\ldots$, és a polinomok váltakozva páros és páratlan függvények.
2. Készítsük el a p_0, p_1, p_2 ortogonális polinomokat a $\{-2,-1,0,1,2\}$ alappontokra!
3. A Csebisev polinomok is ortogonális polinomok, amelyek a következő rekurzióval állíthatók elő: $T_0=1,\ T_1=x,\ T_{n+1}=2xT_n-T_{n-1}$. Ez a szokásos alak, bár így nem 1-főegyütthatósak. Állítsuk elő a $4x^2-3x+2$ polinomot Csebisev-polinomok szerint!
4. $P(x)=\sum_{j=0}^{k}(2j+1)T_j(x)$. Az x_0 helyen ekkor mi a célszerű kiszámítási módja $P(x_0)$-nak?
5. Mutassuk meg, hogy $(p_i,p_i)=\mu_0\beta_1\beta_2\ldots\beta_i$, ahol $\mu_0=(p_0,p_0)\left[=\int_a^b\alpha(x)dx\right]$ a 0-adik momentum.
6. Mutassuk meg, hogy a

$$\begin{pmatrix} x-\alpha_1 & -\beta_1 & & \\ -\beta_1 & x-\alpha_2 & \ddots & \\ & \ddots & \ddots & -\beta_{n-1} \\ & & -\beta_{n-1} & x-\alpha_n \end{pmatrix}$$

háromátlójú mátrix bal felső sarok-aldeterminánsai α_j és β_j^2 parméterekkel bíró ortogonális polinomokat adnak.

10. Lineáris egyenletrendszerek megoldása iterációval

A lineáris egyenletrendszerek megoldását nem mindig célszerű véges módszerrel készíteni. Ha a mátrix nagyméretű és ritka – azaz soronként csak kevés nemzérus elem található – akkor az *LU*-felbontás hátránya, hogy felbontáskor a mátrix nemzérus elemeinek száma megnő – besűrűsödik – ez egyrészt tárolási kérdéseket vet fel, másrészt a sok nemzérus elem miatt megnő a munkaigény. Az iterációs módszereknél ilyen nehézségek nem lépnek fel, de probléma, ha a konvergencia lassú.

10.1. Egyszerű iteráció

Legyen $A \in \mathbb{R}^{n\times n}$ egy felbontása

$$A = M - N \tag{10.1}$$

Ha M invertálható, akkor a következő iterációs módszert készíthetjük: $Ax = (M - N)x = b \rightarrow x = M^{-1}(b + Nx)$, azaz

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b = Bx^{(k)} + c, \tag{10.2}$$

ahol k a vektor felső indexében az iterációszámot jelöli. A B mátrixot *iterációs mátrixnak* nevezzük. Kérdés, mikor remélhető, hogy a fenti iteráció konvergens és az milyen sebességű?

10.1.1 A konvergencia vizsgálata

Az $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ függvény kontrakció, ha van olyan $0 \le q < 1$ szám, amelyre $\forall x, y \in \mathbb{R}^n$ esetén

$$\|F(x) - F(y)\| \le q\|x - y\| \tag{10.3}$$

teljesül. Itt q a *kontrakciós állandó*, vagy *kontrakciószám.* Figyeljük meg, $q < 1$ azt jelenti, hogy a leképezett vektorok közelebb kerülnek egymáshoz.

10.1.2 Tétel. (Banach fixponttétel)

Legyen $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ egy leképezés $q < 1$ kontrakciós állandóval. Akkor

1) $\exists x^* \in \mathbb{R}^n: \ x^* = F(x^*)$, azaz van az iterációnak fixpontja és ez egyértelmű.

2) $\forall x^{(0)} \in \mathbb{R}^n$ kezdőértékre $x^{(k+1)} = F(x^{(k)})$ konvergens sorozat és $\lim\limits_{k\to\infty} x^{(k)} \rightarrow x^*$.

3) Fennáll a hibabecslés: $\left\|x^{(k)} - x^*\right\| \le \dfrac{q^k}{1-q}\left\|x^{(1)} - x^{(0)}\right\|$.

Bizonyítás. Az $x^{(k+1)} = F(x^{(k)})$ sorozat Cauchy-sorozat: $\left\|x^{(k+1)} - x^{(k)}\right\| = \left\|F(x^{(k)}) - F(x^{(k-1)})\right\| \le$ $\le q\left\|x^{(k)} - x^{(k-1)}\right\| \le q^2\left\|x^{(k-1)} - x^{(k-2)}\right\| \le \ldots \le q^k\left\|x^{(1)} - x^{(0)}\right\|$. Így a sorozat egymás utáni tagjai egyre közelebb vannak egymáshoz: van határérték. Legyen most $m \ge k \ge 1$. Ekkor a teleszkópikus összeg képzésével

$$\left\|x^{(m)} - x^{(k)}\right\| = \left\|x^{(m)} - x^{(m-1)} + x^{(m-1)} - x^{(m-2)} + \ldots + x^{(k+1)} - x^{(k)}\right\| \le (q^{m-1} + q^{m-2} + \ldots + q^k)\left\|x^{(1)} - x^{(0)}\right\| =$$

$$= \frac{q^k - q^m}{1-q}\left\|x^{(1)} - x^{(0)}\right\| < \frac{q^k}{1-q}\left\|x^{(1)} - x^{(0)}\right\|.$$

$m \rightarrow \infty$ mellett kapjuk a 3) állítást.

Az egyértelműség igazolásához indirekt módon tegyük fel: van két fixpont, x_1^* és x_2^*. De ekkor a kontrakció felhasználásával $\|x_1^* - x_2^*\| = \|F(x_1^*) - F(x_2^*)\| \le q\|x_1^* - x_2^*\|$, $q < 1$, ami ellentmondás, hiszen a különbség normája nem lehet önmagánál kisebb. ■

A tétel szerint az $x^{(k+1)} = Bx^{(k)} + c$ iteráció konvergens, ha az $F(x) = Bx + c$ leképezés kontrakció:

$$\|F(x) - F(y)\| = \|Bx + c - By - c\| = \|B(x - y)\| \le \|B\|\|x - y\|$$

Innen látható, kontrakció van, ha $\|B\| < 1$. Bizonyítás nélkül megjegyezzük: a spektrál sugár az indukált normák infimuma. Emiatt mondhatjuk: $Bx + c$ konvergens, ha B spektrál sugara $\rho(B) < 1$.
A (10.1) felbontást *regulárisnak* nevezzük, ha M invertálható és $\rho(M^{-1}N) < 1$.

10.2. Jacobi-iteráció

Legyen A felbontása $A = L + D + U$, ahol $D = \mathrm{diag}(A)$, L a mátrix főátló alatti, U a főátló feletti része (szigorúan *alsó* ill. *felső* Δ-mátrixok).
A Jacobi-iterációnál $M = D,\ N = -L - U$ a választás, ezzel

$$B_J = -D^{-1}(L + U),\ \ c_J = D^{-1}b. \tag{10.4}$$

Komponensenkénti alak: $x_i^{(k+1)} = -\dfrac{1}{a_{ii}}\left(\sum_{\substack{j=1\\ j\ne i}}^{n} a_{ij}x_j^{(k)} - b_i\right)$.

Tárigény: $A, b, x^{(k)}, x^{(k+1)}$. Célszerű kezdővektor, ha nincs jobb: $x^{(0)} = c_J$.

10.2.1 Tétel

Ha A sor szerint szigorúan főátló-domináns, akkor a Jacobi-iteráció konvergens.

Bizonyítás. $\|B_J\|_\infty = \max\limits_{(k)}\left\|e_k^T D^{-1}(L + U)\right\|_\infty = \max\limits_{(k)}\sum_{j\ne k}\left|\dfrac{a_{kj}}{a_{kk}}\right| < 1$, tehát van kontrakció.

10.3. Gauss-Seidel iteráció

A Gauss-Seidel iterációnál $M = L + D,\ N = -U$ a választás, ezzel

$$B_{GS} = -(L + D)^{-1}U,\ \ c_{GS} = (L + D)^{-1}b. \tag{10.5}$$

A Gauss-Seidel iteráció komponensenkénti alakját $(L + D)x^{(k+1)} = -Ux^{(k)} + b$ i-edik sorának kiírásából kapjuk:

$$x_i^{(k+1)} = -\frac{1}{a_{ii}}\left(\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} + \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} - b_i\right),\ \ i = 1, 2, \ldots, n. \tag{10.6}$$

Olyan a műveletek sorrendje, hogy $x_i^{(k)}$ értéke $x_i^{(k+1)}$ értékével felülírható. Így a tárigény: A, b, x előnyösebb, mint a Jacobi-iterációnál. Célszerű kezdővektor: $x^{(0)} = c_{GS}$.

10.3.1 Tétel. Felhasításból származó iterációs mátrix normájának becslése

Legyenek A_1, A_2, D $n \times n$-es valós mátrixok, D diagonálmátrix: $e_i^T De_i = d_i$ és $\left\|e_i^T A_1\right\|_\infty < |d_i|$ $\forall$ i-re. Ekkor fennáll:

$$\left\|(A_1+D)^{-1}A_2\right\|_\infty \le \max_{(i)} \frac{\left\|e_i^T A_2\right\|_\infty}{|d_i| - \left\|e_i^T A_1\right\|_\infty},$$

ahol a maximum-keresésnél elegendő a nemzérus számlálókat tekinteni.

Bizonyítás. Mivel $A_1 + D$ főátlódomináns, a kijelölt inverz létezik. A norma definiciója szerint $\left\|(A_1+D)^{-1}A_2\right\|_\infty = \max_{\|x\|_\infty=1}\left\|(A_1+D)^{-1}A_2 x\right\|_\infty$. Legyen $y=(A_1+D)^{-1}A_2 x$ és tegyük fel, a maximum y i-edik indexénél valósul meg: $\|y\|_\infty = |y_i|$. Átrendezéssel $A_2 x = (A_1+D)y \rightarrow Dy = A_2 x - A_1 y$, ahonnan az i-edik sorra $\left\|e_i^T Dy\right\|_\infty = |d_i y_i| \le \left\|e_i^T A_2\right\|_\infty \|x\|_\infty + \left\|e_i^T A_1\right\|_\infty |y_i|$. Figyelembe véve, hogy $\|x\|_\infty = 1$, innen átrendezéssel kapjuk az egyenlőtlenséget. Mivel nem tudjuk, melyik i-re valósul meg $\|y\|_\infty$, ezért a törtet a sorok szerint maximalizáljuk. Ha A_2 i-edik sora zérus, akkor $|d_i||y_i| \le \left\|e_i^T A_1\right\||y_i|$ adódik, ami feltevésünkkel ellentmondó nemzérus $|y_i|$ mellett, így ezeket a sorokat elhagyhatjuk. ■

10.3.2 Tétel

Ha A sor szerint szigorúan domináns átlójú, akkor

$$\|B_{GS}\|_\infty \le \max_{(i)} \frac{\sum_{j=i+1}^{n}|a_{ij}|}{|a_{ii}| - \sum_{j=1}^{i-1}|a_{ij}|} \le \|B_J\|_\infty < 1. \tag{10.7}$$

Bizonyítás. Az előző tételben legyen $A_1 = L$, $U = A_2$ és D a mátrix főátlójából alkotott diagonálmátrix. Ezzel a választással közvetlenül adódik az első egyenlőtlenség.

A második egyenlőtlenség igazolásához vezessük be az

$$\alpha_i = \frac{1}{|a_{ii}|}\sum_{j=1}^{i-1}|a_{ij}|, \quad \text{és} \quad \beta_i = \frac{1}{|a_{ii}|}\sum_{j=i+1}^{n}|a_{ij}| \tag{10.8}$$

jelöléseket. Eszerint $\|B_{GS}\|_\infty \le \max_{(i)} \frac{\beta_i}{1-\alpha_i}$ és igazolandó, hogy ez nem nagyobb, mint $\|B_J\|_\infty = \max_{(j)}(\alpha_j + \beta_j)$. Tegyük fel, $\alpha_j > 0$. Ekkor $(\alpha_j+\beta_j) > \beta_j/(1-\alpha_j)$-ből átrendezéssel $\alpha_j + \beta_j - \alpha_j^2 - \beta_j\alpha_j > \beta_j$, ahonnan $1-\alpha_j-\beta_j > 0$ következik. Ez éppen a szigorú főátló-dominancia feltétele, amit feltettünk. Egyenlőség csak akkor lehetséges, ha $\alpha_j = 0$. Így

$$\|B_{GS}\|_\infty \le \max_{(j)} \frac{\beta_j}{1-\alpha_j} = \frac{\beta_k}{1-\alpha_k} \le \alpha_k + \beta_k \le \max_{(j)}(\alpha_j+\beta_j) = \|B_J\|_\infty.$$

Ha balról az első maximumnál $\alpha_k > 0$, akkor biztosan $\|B_{GS}\|_\infty < \|B_J\|_\infty$. ■

10.4. Gauss-Seidel (GS-) relaxáció

Ekkor a gyorsabb konvergencia reményében D szerepét megosztjuk L és U között:

$$(L+D)x = -Ux + b \qquad \text{/ szorozzuk } \omega\text{-val}$$

$$Dx = Dx \qquad \text{/ szorozzuk } (1-\omega)\text{-val}$$

Összeadva, majd x-re rendezve:

$$(D+\omega L)x^{(k+1)} = (1-\omega)Dx^{(k)} - \omega U x^{(k)} + \omega b \tag{10.9}$$

$$x^{(k+1)} = (D+\omega L)^{-1}\left[(1-\omega)D - \omega U\right]x^{(k)} + (D+\omega L)^{-1}\omega b.$$

Innen az iterációs mátrix

$$B_{GS}(\omega) = (D+\omega L)^{-1}\left[(1-\omega)D - \omega U\right]. \tag{10.10}$$

Ha $\omega = 1$, akkor visszakapjuk a Gauss-Seidel iterációt. A komponensenkénti alakot (10.9) i-edik sorából kapjuk:

$$x_i^{(k+1)} = -\frac{\omega}{a_{ii}}\left(\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} + \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} - b_i\right) + (1-\omega)x_i^{(k)},\quad i = 1,2,\ldots,n. \tag{10.11}$$

Eszerint a Gauss-Seidel relaxáció következő lépésének eredményét megszorozzuk ω-val és ehhez hozzáadjuk a k-adik vektor $(1-\omega)$-szorosát.

10.5. A relaxációs módszerekre vonatkozó néhány tétel

1. Ha A egyik diagonáleleme sem 0, egyébként tetszőleges, akkor $\rho(B_{GS}(\omega)) \geq |\omega - 1|$, azaz csak akkor remélhető konvergencia, ha ω 0 és 2 közé esik.
2. Legyen $A \in \mathbb{R}^{n\times n}$ szimmetrikus, pozitív definit mátrix és $0 < \omega < 2$. Ekkor $\rho(B_{GS}(\omega)) < 1$, vagyis minden ilyen ω-ra konvergens a GS-relaxáció.

A következő két tétel blokk-háromátlós mátrixokra vonatkozik. Természetesen 1×1-es blokkok esetén a megszokott mátrixot kapjuk vissza.

3. Legyen $A \in \mathbb{R}^{n\times n}$ blokk-háromátlós mátrix. Akkor a megfelelő blokk Jacobi (J) és GS-iteráció mátrixaira

   $$\rho(B_{GS}^b) = \left[\rho(B_J^b)\right]^2.$$

   Ez azt jelenti, a kettő egyszerre konvergens, vagy divergens és konvergencia esetén a GS-iteráció kétszer gyorsabb.
4. Legyen A blokk-háromátlós, szimmetrikus és pozitív definit. Ekkor a blokk-Jacobi iteráció, valamint a blokk-GS relaxáció $0 < \omega < 2$ mellett konvergens. Utóbbinál az optimális relaxációs paraméter

   $$\omega_0 = 2/\left(1 + \sqrt{1 - \left(\rho(B_J^b)\right)^2}\right) \in (0,2)$$

és erre az optimális paraméterre a spektrál sugár

$$\rho(B_{GS}^b(\omega_0)) = |\omega_0 - 1| < \rho(B_{GS}^b) = \left(\rho(B_J^b)\right)^2.$$

10.6. Egy lépésben optimális ω paraméter meghatározása

Láttuk, (10.11)-ben az x_k vektorból kiindulva az

$$x_{k+1}^{\omega} = \omega x_{k+1} + (1-\omega)x_k = x_k + \omega(x_{k+1} - x_k) \tag{10.12}$$

vektort készítjük, a Gauss-Seidel módszer x_{k+1} vektora helyett. Vezessük be az $y_k = x_{k+1} - x_k$, $r_k = b - Ax_k$ jelöléseket és határozzuk meg az ω paramétert általánosan az $A = M - N$ felbontás mellett! (10.12)-ból kapjuk:

$$r_{k+1} = b - Ax_{k+1}^{\omega} = r_k - \omega A y_k \,. \tag{10.13}$$

Határozzuk meg a k -adik lépésben ω_k -t abból a feltételből, hogy $\|r_{k+1}\|_2$ minimális! Ehhez nem kell mást tenni, mint az $Ay_k\omega = r_k$ „egyenletet” a pszeudoinverzzel ω -ra megoldani:

$$\omega_k = \left(Ay_k\right)^+ r_k = \frac{y_k^T A^T r_k}{\|Ay_k\|_2^2} = \frac{r_k^T A y_k}{\|Ay_k\|_2^2} \tag{10.14}$$

Hogy ne kelljen x_{k+1} -et explicit módon előállítani, a relaxáció nélküli alakból kifejezzük y_k -t:

$$x_{k+1} = M^{-1}(Nx_k + b) = x_k + M^{-1}\left(b - (M - N)x_k\right) = x_k + M^{-1} r_k \,, \tag{10.15}$$

innen

$$y_k = M^{-1} r_k \,. \tag{10.16}$$

Az ω_k meghatározásához vezessünk be egy újabb vektort:

$$c_k = Ay_k = (M - N)M^{-1} r_k = r_k - N y_k \,, \tag{10.17}$$

és ekkor a következő iterációs algoritmust készíthetjük:

Kezdés: $r_0 = b - Ax_0$;

$k = 1, 2, 3, \ldots,$-ra:

$y_k = M^{-1} r_k$;

$c_k = r_k - N y_k$;

$\omega_k = \dfrac{r_k^T c_k}{c_k^T c_k}$;

$x_{k+1} = x_k + \omega_k * y_k$;

$r_{k+1} = r_k - \omega_k * c_k \quad (= b - Ax_{k+1})$;

Két lehetőség is van r_{k+1} számítására. Természetesen az első az olcsóbb. Az iteráció előrehaladtával lehet, hogy a második módszer eredménye jelentősen eltér az elsőétől. Célszerű ilyenkor r_{k+1} értékét a második, pontosnak tekinthető módszerrel feljavítani. Az algoritmusban a vektorokat indexeltük, bár nem szükséges, mivel minden lépésben az előző vektor az újjal felülírható.

10.7. A Richardson iteráció

Ha a mátrixunk sajátértékei valós, pozitív számok, akkor egy iterációt készíthetünk a következő észrevétel alapján:

$$(I - pA)r_i = r_{i+1} = b - A(x_i + pr_i) = b - Ax_{i+1}, \quad x_{i+1} = x_i + pr_i \,, \tag{10.18}$$

ahol a p számot úgy választjuk, hogy $I - pA$ spektrál sugara minél kisebb legyen. Az $I - pA$ mátrix sajátértékei most $1 - p\lambda_i$-k. Látjuk, az $1 - px$ leképező függvény a (0,1) ponton áthaladó, pozitív p mellett negatív meredekségű egyenes. Legyen a legkisebb sajátérték m, a legnagyobb M. A Richarson iterációnál az optimális p értéket abból a feltételből határozzuk meg, hogy a legkisebb és a legnagyobb sajátérték ugyanakkora abszolút értékű számba képződjön le:

$$1 - pm = -(1 - pM) \quad \to \quad p = 2/(m + M). \tag{10.19}$$

Ezzel a választással $I - pA$ spektrál sugara $(M - m)/(M + m)$ lesz.

H nem ismerjük a mátrix sajátértékeit, de tudjuk, hogy a sajátértékek pozitívak, pl. mert A szimmetrikus, pozitív definit, a p számot abból a feltételből is kereshetjük, hogy $\|r_{i+1}\|_2$ legyen minimális. Ekkor az $r_i = pAr_i$ egyenlet pszeudo-megoldása

$$p = \frac{r_i^T A^T r_i}{\|Ar_i\|_2^2} = \frac{r_i^T A r_i}{\|Ar_i\|_2^2}. \tag{10.20}$$

Az iteráció során elég néhányszor kiszámolni p értékét, hiszen az az előbb megállapított optimális érték körül fog ingadozni.

10.8. Feladatok

1. Bizonyítsuk be, szimmetrikus mátrixokra a Rayleigh-hányados legkisebb értéke a legkisebb sajátérték.
2. Hogy hajtsuk végre a Jacobi-iterációt, ha a mátrix az oszlopai szerint szigorúan főátló-domináns?
3. Mutassuk meg, a 10.3.1 tétel átfogalmazható arra az esetre, amikor a mátrix oszlopai szerint szigorúan főátló-domináns.
4. Mit ad a (10.7) becslés arra az esetre, ha a sor szerinti főátló-dominancia csak úgy teljesül, hogy néhány sorban van egyenlőség? És ha csak az utolsó sorban van egyenlőség?
5. $A = \begin{pmatrix} 5 & -1 & 2 & 1 \\ -3 & 7 & -2 & 0 \\ 3 & 0 & 5 & -1 \\ 0 & 2 & -4 & 6 \end{pmatrix}$. $\|B_J\|_\infty = ?$ $\|B_{GS}\|_\infty \le ?$
6. A 10.3.1 Tétel segítségével bizonyítsuk be: $\|A^{-1}\|_\infty \le \max_i \frac{1}{|a_{ii}|(1-\alpha_i-\beta_i)}$, ld. (10.8)-at is, ha A sorai szerint szigorúan főátló-domináns. Oszlopok szerinti főátló-dominancia esetén hogyan módosítsuk az állítást?
7. Feltéve, hogy $D+\omega L$ főátlódomináns a sorai szerint, a 10.3.1 Tétel segítségével mutassuk meg: $\|B_{GS}(\omega)\|_\infty \le \max_{(j)} \frac{|1-\omega|+\omega\beta_j}{1-\omega\alpha_j}$.
8. Ha $\|D^{-1}A_1\|<1$, a 10.3.1 Tételhez hasonló egyenlőtlenséget származtathatunk (2.15) felhasználásával, mivel $(A_1+D)^{-1}A_2 = (I+D^{-1}A_1)^{-1}D^{-1}A_2$. Mutassuk meg, hogy ekkor indukált normával érvényes: $\|(A_1+D)^{-1}A_2\| \le \frac{\|D^{-1}A_2\|}{1-\|D^{-1}A_1\|}$. Szükséges, hogy most D diagonálmátrix legyen? Az 5. Példa mátrixára melyik módszer ad jobb becslést?