Tartalomjegyzék

1. Bevezetés

Dr. Abaffy József tanár úr három témát hirdetett meg Charles George Broyden barátja emlékére a 2017/18/1 félévben. Ezek a következők voltak:

1. Broyden és a feltétel nélküli optimalizálás,
2. Broyden és a nemlineáris egyenletrendszerek megoldása,
3. Broyden és a lineáris egyenletrendszerek megoldása.

Én a harmadik témát választottam. A lineáris algebra szerepe kiemelkedően fontos rengeteg területen, például jelentős geometriai, fizikai és mérnöki alkalmazásokkal rendelkezik, de a modern társadalomtudományokban is alkalmazzák. Számos más területen is találkozhatunk lineáris algebrával, szinte minden tudományág tartalmaz olyan modelleket, amelyek lineáris egyenletrendszerek megoldására vezetik vissza valamilyen probléma megoldását.

Broyden-en kívül természetesen más szerzők műveiből is merítünk (lásd a hivatkozásoknál), de mivel a témakör igen nagy, teljességre nem törekedhettünk e téren. A diplomamunka pl. a legtöbb tétel bizonyítását tartalmazza, de van, ahol csak hivatkozunk a bizonyításokra. Vannak olyan eredmények, amiket a hely és az idő végessége miatt teljesen kihagytunk, de természetesen ezek a kevésbé fontos eredmények.

A diplomamunka Broyden rövid életrajzával kezdődik. A 3. és a 4. fejezet a további fejezeteket alapozzák meg. Ezekben a fejezetekben a további fejezetekhez szükséges fogalmakat vezetünk be. Definíciókat adunk meg és tételeket mondunk ki. Tisztázzuk a jelöléseket. A témakör felvezetése annyiban különleges, hogy egyáltalán nem használjuk a determináns fogalmát. Ebben Broyden-hez igazodunk, mivel ezek a fejezetek nagyban támaszkodnak a Basic Matrices [3] című könyvére. A fejezetek struktúrája szintén kicsit eltér a megszokottól. Ebben szintén Broyden-hez igazodtunk. A témakör más irányból való megközelítése számomra igen hasznos volt. Az 5-7. fejezetek Broyden, a numerikus lineáris algebra témakörében elért eredményeit foglalják össze. Az 5. fejezetben vizsgáljuk a felső relaxációs eljárás konvergencia kritériumait. A 6. fejezetben egy átfogó rendszertant adunk a konjugált gradiens módszerekre. A 7. fejezetben a blokk konjugált gradiens módszert vizsgáljuk. A 8. fejezetben MATLAB tesztfeladatokon hasonlítjuk össze a pcg, SYMMLQ, bicgstabl, QMR és az LSQR módszereket.

Szeretném megköszönni Dr. Abaffy József segítségét a diplomamunka írásával kapcsolatban.

2. Broyden életrajza

Charles George Broyden (1933. február 3. - 2011. május 20.) szerény családi háttérrel, Angliában született. Édesapja gyári munkásként, édesanyja háztartásbeliként dolgozott. Szülei ennek ellenére a kezdetektől tanulásra biztatták. Charles már gyerekként rengeteget olvasott, az iskolában jól teljesített. Édesapja sajnálatos módon meghalt tuberkulózisban, mikor Charles még csak 11 éves volt. Ez még jobban megnehezítette családi helyzetüket, de édesanyja így is arra biztatta, hogy egyetemre menjen. A King's College London egyetemen szerzett fizikus diplomát 1955-ben. A következő 10 évet az iparban töltötte. Ezután 1965-1967 között az Aberystwyth Egyetemen tanított, majd a University of Essex egyetemen 1967-ben professzor, később a matematika intézet dékánja lett. 1986-ban innen visszavonult, 1990-ben a Bolognai Egyetemen fogadott el professzori kinevezést. Jelentős szerepe volt a kvázi-Newton módszerek kifejlesztésében. A kvázi-Newton módszerek előtt a nemlineáris optimalizálási problémákat gradiens alapú módszerekkel oldották meg. Nemlineáris esetben az ehhez szükséges Hessian mátrix kiszámítása legtöbbször nem praktikus. A kvázi-Newton módszerek kifejlesztésére irányuló munka az 1960-as és 1970-es években zajlott, a nemlineáris optimalizálás egy izgalmas időszakában. A kutatásban részt vett még például Bill Davidon, Roger Flecher és Mike Powell. A kutatások eredménye valódi ipari alkalmazások problémáinak megoldására adott eszközöket.

Az iparban töltött évei alatt Broyden a Davidon-Fletcher-Powell (DFP) módszert adaptálta nemlineáris problémákra. Ez vezetett az 1965-ös klasszikus "A class of methods for solving nonlinear simultaneous equations" cikkéhez a Mathematics of Computation folyóiratban. Ez a munka széleskörűen elismert, mint a 20. század egyik legnagyobb numerikus analízis eredménye. A University of Essex egyetemen a DFP módszer optimalizálásra fókuszált. Feltűnt neki, hogy habár a módszer jól működik, néha furcsa eredményeket produkál. Kerekítési hibákra gyanakodott. A kutatása 1970-ben egy új, továbbfejlesztett módszerhez vezetett. Tőle függetlenül, nagyjából egy időben, Fletcher, Donald Goldfarb, és David Shanno is ugyanerre az eredményre jutott. Ezért az új módszert a neveik kezdőbetűiből BFGS módszernek

nevezték el. Más kutatások folytatták a kvázi-Newton módszerek optimalizálását, de a BFGS módszer még ma is a leginkább választott, ha a Hessian mátrix kiszámítása túl költséges. 1981-től Abaffy Józseffel és Emilio Spedicato-val az ABS (Abaffy-Broyden-Spedicato) módszereken dolgozott.

Később a numerikus lineáris algebrára fókuszált, ezen belül is a konjugált gradiens módszerekre és ezek rendszertanára. A diplomamunka a kutatásainak a konjugált gradiens módszerekkel kapcsolatos részét dolgozza fel. 2011-ben 78 évesen, egy szívroham komplikációiba belehalt.

Feleségével, Joan-nal, 1959-ben házasodtak össze. Négy gyerekük született, a legidősebb, Robbie, sajnos 4 éves korában meghalt. Broyden nagy örömét lelte családjában, gyermekeiben, Christopher-ben, Jane-ben és Nicholas-ban. Szeretett madárlesre járni, zenével foglalkozni, kórusban énekelni, vitorlázni. A helyi közösség és az egyházi közösség aktív tagja volt. Hét unokája született. A legidősebb unokája, Tom, az Oxford egyetemen tanult matematikát, a második legidősebb unokája, Matt, a Warwick Egyetemen tanul matematikát, Ben pedig mérnöknek tanul a Swansea Egyetemen. Így nagyapjuk nyomában járnak.

Köszönet Joan Broyden-nek a életrajzban nyújtott segítségéért.

3. Mátrixok alapvető jellemzése

Ebben a fejezetben a további fejezetekhez szükséges fogalmakat vezetünk be. Definíciókat adunk meg és tételeket mondunk ki. Tisztázzuk a jelöléseket. [3] [13]

3.1. Vektorok és mátrixok

Definíció. A valós vektor a valós számok egy rendezett halmaza.

Definíció. Egy vektor elemeinek a száma a vektor rendje, vagy más szóval a vektor dimenziója.

A vektorokat a diplomamunkában vastag kisbetűvel jelöljük. Például $\boldsymbol{x} = [x_i]$, ahol x_i a vektor i-edik elemét jelöli. Oszlopvektoron vektort, sorvektoron vektor transzponáltat értünk. Az $\boldsymbol{x}$ vektor transzponáltját $\boldsymbol{x}^T$-vel jelöljük.

Példa. Ha $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, akkor $\boldsymbol{x}^T = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$.

Definíció. Legyenek $\boldsymbol{x} = [x_i]$, $\boldsymbol{y} = [y_i]$ és $\boldsymbol{z} = [z_i]$ n-ed rendű vektorok. Legyen $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$. Ekkor $z_i = x_i + y_i$.

Definíció. Legyen $\boldsymbol{x} = [x_i]$ és $\boldsymbol{y} = [y_i]$ n-ed rendű valós vektor. A belső szorzata, vagy más néven skaláris szorzata a sorvektor $\boldsymbol{x}^T$-nek és az oszlopvektor $\boldsymbol{y}$-nak

$$\boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i \,. \tag{3.1}$$

Definíció. Az $\boldsymbol{x}$ és $\boldsymbol{y}$ vektorok egymásra ortogonálisak, ha a belső szorzatuk 0.

Definíció. Legyenek $\boldsymbol{p}_i$ vektorok és y_i skalárok, $i = 1,2,...,n$. A

$$\sum_{i=1}^{n} \boldsymbol{p}_i y_i \tag{3.2}$$

vektor a $\boldsymbol{p}_i$ vektorok lineáris kombinációja.

Definíció. A valós mátrix azonos rendű valós vektorok egy rendezett halmaza.

A mátrixokat a diplomamunkában vastag nagybetűvel jelöljük. Egy $m\times n$-es mátrixot értelmezhetünk úgy, mint m darab sorvektor, vagy mint n darab oszlopvektor. Röviden azt mondjuk, hogy a mátrix sorai és oszlopai. Az $\boldsymbol{A} = [a_{ij}]$ jelölésben a_{ij} az $\boldsymbol{A}$ mátrix i-edik sorának j-edik eleme. A csupa nullából álló mátrixot vagy vektort $\boldsymbol{0}$-val jelöljük. Egy mátrixra akkor mondjuk, hogy ritka, ha viszonylag kevés nullától különböző eleme van. Egy mátrixra akkor mondjuk, hogy kitöltött, ha viszonylag kevés nulla eleme van.

Példa. Legyen $\boldsymbol{A} = [a_{ij}]$ $m\times n$-es mátrix. Ekkor $i = 1,2,...,m$ és $j = 1,2,...,n$.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix}$$

Definíció. Ha egy mátrixnak ugyanannyi sora és oszlopa van, négyzetes mátrixnak nevezzük.

Definíció. A négyzetes mátrix rendje az oszlopainak száma.

Definíció. Az $\boldsymbol{A} = [a_{ij}]$ mátrix transzponáltját $\boldsymbol{A}^T$-vel jelöljük, jelentése $\boldsymbol{A}^T = [a_{ji}]$, azaz a mátrix sorainak és oszlopainak felcserélésével kapott mátrix.

Definíció. Legyen $\boldsymbol{A} = [a_{ij}]$ mátrix. A mátrix diagonális elemei azok az elemek, ahol $i = j$.

Definíció. Az $n \times n$-es $\boldsymbol{D} = [d_{ij}]$ mátrix diagonális mátrix, ha $d_{ij} = 0$ minden $i \neq j$ indexre. A diagonális mátrixot $diag(d_i)$-vel jelöljük, ahol d_i az i-edik diagonális elem.

Definíció. A $diag(d_i)$ mátrixot, ahol $d_i = 1$ minden i-re, egységmátrixnak nevezzük, és $\boldsymbol{I}$-vel jelöljük.

Definíció. Legyenek $\boldsymbol{A} = [a_{ij}]$, $\boldsymbol{B} = [b_{ij}]$ és $\boldsymbol{C} = [c_{ij}]$ $m \times n$-es mátrixok. Legyen $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$. Ekkor $c_{ij} = a_{ij} + b_{ij}$.

Definíció. Legyen $\boldsymbol{A}$ $m \times n$-es mátrix. Legyen $\boldsymbol{x} = [x_i]$ n-ed rendű vektor, $\boldsymbol{y} = [y_i]$ m-ed rendű vektor. Jelölje $\boldsymbol{a}_i^T$ az $\boldsymbol{A}$ mátrix i-edik sorát. Az $\boldsymbol{A}$ mátrix és $\boldsymbol{x}$ vektor mátrix-vektor szorzata $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ és

$$y_i = \boldsymbol{a}_i^T \boldsymbol{x} \,. \tag{3.3}$$

Definíció. Legyen $\boldsymbol{A}$ $m \times n$-es mátrix és jelöle $\boldsymbol{a}_i^T$ az $\boldsymbol{A}$ mátrix i-edik sorát. Legyen $\boldsymbol{B}$ $n \times p$-s mátrix és jelöle $\boldsymbol{b}_j$ a $\boldsymbol{B}$ mátrix j-edik oszlopát. Az $\boldsymbol{AB}$ mátrixszorzat eredménye a $\boldsymbol{C} = [c_{ij}]$ $m \times p$-s mátrix, ahol

$$c_{ij} = \boldsymbol{a}_i^T \boldsymbol{b}_j \,. \tag{3.4}$$

Az $\boldsymbol{AB}$ mátrixszorzat esetén az $\boldsymbol{A}$ mátrix oszlopainak száma meg kell egyezzen a $\boldsymbol{B}$ mátrix sorainak számával, hogy a megfelelő sorvektorok és oszlopvektorok belső szorzatai jól definiáltak legyenek. A mátrixszorzás általában nem kommutatív, $\boldsymbol{AB} \neq \boldsymbol{BA}$. Egy mátrixot az egységmátrixszal szorozva önmagát kapjuk, $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$.

Definíció. Az $\boldsymbol{A}$ mátrix idempotens, ha $\boldsymbol{A}^2 = \boldsymbol{A}$.

Definíció. Az $\boldsymbol{A}$ mátrix szimmetrikus, ha $\boldsymbol{A} = \boldsymbol{A}^T$. Az $\boldsymbol{A}$ mátrix antiszimmetrikus, ha $\boldsymbol{A} = -\boldsymbol{A}^T$.

Mivel $(\boldsymbol{A}^T)^T = \boldsymbol{A}$ és $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$, ezért $(\boldsymbol{A}^T\boldsymbol{A})^T = \boldsymbol{A}^T(\boldsymbol{A}^T)^T = \boldsymbol{A}^T\boldsymbol{A}$. Azaz az $\boldsymbol{A}^T\boldsymbol{A}$ valós mátrix mindig szimmetrikus.

Definíció. Legyen $\boldsymbol{x}^T = [\boldsymbol{x}_1^T \ \boldsymbol{x}_2^T]$ n-ed rendű vektor, ahol $\boldsymbol{x}_1^T = [x_1, x_2, ..., x_r]$ és $\boldsymbol{x}_2^T = [x_{r+1}, x_{r+2}, ..., x_n]$ és $1 \leq r < n$. Az $\boldsymbol{x}_1$ és $\boldsymbol{x}_2$ vektorok az $\boldsymbol{x}$ vektor részvektorai, vagy partíciói.

Példa. Legyenek $\boldsymbol{x} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2]$, $\boldsymbol{y} = [\boldsymbol{y}_1 \ \boldsymbol{y}_2]$, $\boldsymbol{u}$ n-ed rendű vektorok. Legyen $\boldsymbol{u} = \boldsymbol{x} + \boldsymbol{y}$. Ha $\boldsymbol{x}_1 = [x_1, x_2, ..., x_r]$, $\boldsymbol{x}_2 = [x_{r+1}, x_{r+2}, ..., x_n]$, $\boldsymbol{y}_1 = [y_1, y_2, ..., y_r]$, $\boldsymbol{y}_2 = [y_{r+1}, y_{r+2}, ..., y_n]$, $1 \leq r < n$, akkor $\boldsymbol{u} = [\ \boldsymbol{x}_1 + \boldsymbol{x}_2 \quad \boldsymbol{y}_1 + \boldsymbol{y}_2\ ]$. Igaz az is, hogy $\boldsymbol{x}^T\boldsymbol{y} = \boldsymbol{x}_1^T\boldsymbol{y}_1 + \boldsymbol{x}_2^T\boldsymbol{y}_2$.

Mátrixokat is részmátrixokra particionálhatunk. Ennek nagy jelentősége, hogy nagy mátrixokat egyszerűbben kezelhetünk.

Definíció. Legyen $\boldsymbol{A}$ $m\times n$-es mátrix. Az $\boldsymbol{A}$ mátrixból annak k számú sora ($1 \leq k \leq$ $\leq m-1$) és l számú oszlopa ($1 \leq l \leq n-1$) törlésével előállított $(m-k)\times(n-l)$-es részmátrixot az $\boldsymbol{A}$ mátrix minormátrixának nevezzük.

A művelettartás a minormátrixokra két particionált mátrix között nem feltétlen jól definiált.

Definíció. Particionált mátrixok egy halmaza egy művelet szerint jól particionált, ha a mátrixok minormátrixai között a művelet jól definiált.

Példa. Az $\boldsymbol{E}_1$ és $\boldsymbol{F}_1$ mátrixok az összeadás szerint jól particionáltak.

$$\boldsymbol{E}_1 = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_{11} \\ a_{21} & a_{22} & a_{23} & b_{21} \\ \hline c_{11} & c_{12} & c_{13} & d_{11} \\ c_{21} & c_{22} & c_{23} & d_{21} \end{array}, \quad \boldsymbol{F}_1 = \begin{bmatrix} \boldsymbol{J} & \boldsymbol{K} \\ \boldsymbol{L} & \boldsymbol{M} \end{bmatrix} = \begin{array}{ccc|c} j_{11} & j_{12} & j_{13} & k_{11} \\ j_{21} & j_{22} & j_{23} & k_{21} \\ \hline l_{11} & l_{12} & l_{13} & m_{11} \\ l_{21} & l_{22} & l_{23} & m_{21} \end{array}$$

A szorzás szerint $\boldsymbol{E}_1$ és $\boldsymbol{F}_1$ nem jól particionáltak, mert a megfelelő minormátrixok szorzata (3.4) szerint nem képezhető.

Példa. Az $\boldsymbol{E}_2$ és $\boldsymbol{F}_2$ mátrixok a szorzás szerint jól particionáltak.

$$\boldsymbol{E}_2 = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_{11} \\ a_{21} & a_{22} & a_{23} & b_{21} \\ \hline c_{11} & c_{12} & c_{13} & d_{11} \\ c_{21} & c_{22} & c_{23} & d_{21} \end{array}, \quad \boldsymbol{F}_2 = \begin{bmatrix} \boldsymbol{J} & \boldsymbol{K} \\ \boldsymbol{L} & \boldsymbol{M} \end{bmatrix} = \begin{array}{c|ccc} j_{11} & k_{11} & k_{12} & k_{13} \\ j_{21} & k_{21} & k_{22} & k_{23} \\ j_{31} & k_{31} & k_{32} & k_{33} \\ \hline l_{11} & m_{11} & m_{12} & m_{13} \end{array}$$

Azonban az összeadás szerint $\boldsymbol{E}_2$ és $\boldsymbol{F}_2$ nem jól particionáltak, mert a megfelelő minormátrixok összege nem képezhető.

Definíció. Legyen $\boldsymbol{A}$ $n\times n$-es mátrix. Legyen $\boldsymbol{A}_{11}$ az $\boldsymbol{A}$ mátrix egy minormátrixa. Ha az $\boldsymbol{A}_{11}$ minormátrix elemei az $\boldsymbol{A}$ mátrix főátlójára szimmetrikusan helyezkednek el, az $\boldsymbol{A}_{11}$ minormátrixot főminormátrixnak nevezzük.

Definíció. Legyen $\boldsymbol{A}$ $n\times n$-es mátrix. Legyen $\boldsymbol{A}$ egy particionálása

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix},$$

ahol $\boldsymbol{A}_{11}$ $r\times r$-es ($r < n$) főminormátrix. Ekkor $\boldsymbol{A}_{11}$ az $\boldsymbol{A}$ mátrix bal felső r-ed rendű sarokminormátrixa.

Példa. Legyen

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Az $\boldsymbol{A}$ mátrix bal felső elsőrendű sarokminormátrixa $[a_{11}]$, a bal felső másodrendű sarokminormátrixa

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

és a bal felső harmadrendű sarokminormátrixa önmaga.

Az alkalmazásokban gyakran szükség van komplex vektorokra, mátrixokra. A belső szorzatot leszámítva a fenti definíciók érvényesek komplex vektorokra és mátrixokra is, azzal a különbséggel, hogy a komplex vektorok elemei komplex számok, a komplex mátrixok elemei komplex vektorok. A belső szorzat általánosabb definíciója megköveteli, hogy egy vektor saját magával vett belső szorzata valós, nem negatív, és csak akkor nulla, ha a vektor nulla. Definiáljuk a belső szorzatot komplex vektorokra is.

Definíció. A $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ komplex vektor konjugáltja $\overline{\boldsymbol{z}} = \boldsymbol{x} - i\boldsymbol{y}$.

Definíció. Az $\boldsymbol{A} = [a_{ij}]$ komplex mátrix konjugáltja $\overline{\boldsymbol{A}} = [\overline{a_{ij}}]$.

Definíció. A $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ komplex vektor Hermite-féle transzponáltja

$$\boldsymbol{z}^H = \overline{\boldsymbol{z}}^T = \boldsymbol{x}^T - i\boldsymbol{y}^T. \tag{3.5}$$

Definíció. Az $\boldsymbol{A} = \boldsymbol{B} + i\boldsymbol{C}$ komplex mátrix Hermite-féle transzponáltja

$$\boldsymbol{A}^H = \overline{\boldsymbol{A}}^T = \boldsymbol{B}^T - i\boldsymbol{C}^T. \tag{3.6}$$

Definíció. Az $\boldsymbol{A}$ komplex mátrix Hermite-mátrix, ha $\boldsymbol{A} = \boldsymbol{A}^H$.

Példa. Legyen $\boldsymbol{A} = \boldsymbol{B} + i\boldsymbol{C}$ és $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$. Ekkor $\boldsymbol{A}^H\boldsymbol{z} = (\boldsymbol{B}^T - i\boldsymbol{C}^T)(\boldsymbol{x} + i\boldsymbol{y}) = = \boldsymbol{B}^T\boldsymbol{x} + \boldsymbol{C}^T\boldsymbol{y} + i(\boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{C}^T\boldsymbol{x})$.

A komplex Hermite-mátrixot így a valós szimmetrikus mátrix analógiájára definiáltuk. Hasonlóan a valós szimmetrikus esethez, a komplex esetre is igaz, hogy $(\boldsymbol{A}^H\boldsymbol{A})^H = \boldsymbol{A}^H\boldsymbol{A}$.

Definíció. A $\boldsymbol{w}$ komplex vektor és a $\boldsymbol{z}$ komplex vektor belső szorzata $\boldsymbol{z}^H\boldsymbol{w}$.

A $\boldsymbol{z}^H\boldsymbol{z} = (\boldsymbol{x}^T - i\boldsymbol{y}^T)(\boldsymbol{x} + i\boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{y}^T\boldsymbol{y}$ szorzat nem lehet se komplex, se negatív, és csak akkor nulla, ha $\boldsymbol{z}$ vektor nulla, így a definíció eleget tesz az általánosabb feltételeknek.

3.2. Lineáris függetlenség

A lineáris függetlenség alapvető fogalom. A következő fejezetekben nagyban fogunk annak a következményeire támaszkodni, hogy vektorok lineárisan függetlenek-e, vagy sem.

Definíció. Az $\boldsymbol{a}_i$, $i = 1,2,...,n$ vektorok lineárisan összefüggők, ha léteznek olyan x_i skalárok, amelyek nem mindegyike zérus, és a velük képzett lineáris kombinációra fennáll, hogy

$$\sum_{i=1}^{n} \boldsymbol{a}_i x_i = \mathbf{0}. \tag{3.7}$$

Ellenkező esetben az $\boldsymbol{a}_i$ vektorok lineárisan függetlenek.

Példa. Az $\boldsymbol{a}_1^T = [\ 1 \quad 2 \quad -1\ ]$, $\boldsymbol{a}_2^T = [\ -2 \quad -1 \quad 1\ ]$, $\boldsymbol{a}_3^T = [\ -1 \quad 4 \quad -1\ ]$ vektorok lineárisan összefüggők, mert $3\boldsymbol{a}_1 + 2\boldsymbol{a}_2 - \boldsymbol{a}_3 = \mathbf{0}$.

Példa. Az $\boldsymbol{a}_1^T = [\ 1 \quad 0 \quad 0\ ]$, $\boldsymbol{a}_2^T = [\ 1 \quad 1 \quad 0\ ]$, $\boldsymbol{a}_3^T = [\ -1 \quad 1 \quad 1\ ]$ vektorok lineárisan függetlenek, mert

$$\sum_{i=1}^{n} \boldsymbol{a}_i x_i = \begin{bmatrix} x_1 + x_2 - x_3 \\ x_2 + x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Ebből következik, hogy $x_3 = 0$. Így $x_2 + x_3 = 0$ és $x_1 + x_2 + x_3 = 0$ akkor és csak akkor, ha x_1 és x_2 is nulla.

Definíció. Az $\boldsymbol{A}$ $m \times n$-es mátrix oszlopai lineárisan összefüggők, ha létezik olyan n-ed rendű $\boldsymbol{x} \neq \mathbf{0}$ vektor, hogy $\boldsymbol{Ax} = \mathbf{0}$. Ha nem létezik ilyen vektor, akkor a mátrix oszlopai lineárisan függetlenek.

Ha az $\boldsymbol{A}$ mátrix oszlopai lineárisan összefüggők, abból nem csak az következik, hogy az oszlopok megfelelő lineáris kombinációja nulla, hanem hogy létezik olyan $\boldsymbol{x} \neq \mathbf{0}$ vektor, amely az $\boldsymbol{A}$ mátrix minden sorára ortogonális. Az $\boldsymbol{x}$ vektor nem egyértelmű, mert a skalárral szorzása nem változtat az ortogonalitáson, így $\boldsymbol{x}$ tetszőlegesen méretezhető. Az $\boldsymbol{A}$ mátrix oszlopainak lineáris függetlenségét nem befolyásolja, ha az $\boldsymbol{A}$ mátrix sorait felcseréljük.

Definíció. A lineárisan független n-ed rendű vektorok halmazát, amelyből lineáris kombinációként bármely más n-ed rendű vektor kifejezhető, bázisnak nevezzük.

3.1. Tétel. *Az $n + 1$ darab n-ed rendű vektor lineárisan összefüggő. [3, 29. oldal, Theorem 2.2]*

Bizonyítás. Indukcióval bizonyítunk. Megmutatjuk, hogy ha bármely $(r-1)\times r$-es mátrix oszlopai lineárisan összefüggők, akkor bármely $r\times(r+1)$-es mátrix oszlopai is lineárisan összefüggők. Legyen $\boldsymbol{A}_1$ $r\times(r+1)$-es mátrix és

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{c}^T & \delta \end{bmatrix},$$

ahol $\boldsymbol{A}$ $(r-1)\times r$-es mátrix, $\boldsymbol{b}$ oszlopvektor, $\boldsymbol{c}^T$ sorvektor és δ skalár. Az $\boldsymbol{A}-\boldsymbol{b}\delta^{-1}\boldsymbol{c}^T$ egy $(r-1)\times r$-es mátrix, és az indukciós feltétel szerint bármely $(r-1)\times r$-es mátrix oszlopai lineárisan összefüggők. Ezért létezik egy olyan $\boldsymbol{x}\neq\boldsymbol{0}$ vektor, hogy

$$(\boldsymbol{A}-\boldsymbol{b}\delta^{-1}\boldsymbol{c}^T)\boldsymbol{x}=\boldsymbol{0}.$$

Ebből következik, hogy

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{c}^T & \delta \end{bmatrix}\begin{bmatrix} \boldsymbol{x} \\ \delta^{-1}\boldsymbol{c}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}.$$

Mivel $\boldsymbol{x}\neq\boldsymbol{0}$, léteznie kell egy olyan $\boldsymbol{y}\neq\boldsymbol{0}$ vektornak, hogy $\boldsymbol{A}_1\boldsymbol{y}=\boldsymbol{0}$. Tehát $\boldsymbol{A}_1$ sorai is lineárisan összefüggők. □

Következmény. *Az n-ed rendű vektorok bázisa mindig n darab vektor.*

3.2. Tétel. *Legyen r darab lineárisan független vektorunk, amik egy új vektor hozzáadásával lineárisan összefüggővé válnak. Ekkor az új vektor kifejezhető az eredeti r vektorok egyértelmű lineáris kombinációjaként. [3, 30. oldal, Lemma 2.2]*

Bizonyítás. Legyen $\boldsymbol{A}$ $n\times r$-es mátrix és $\boldsymbol{b}$ n-ed rendű vektor. Legyenek az $\boldsymbol{A}$ mátrix oszlopai lineárisan függetlenek. Legyen a $\boldsymbol{b}$ vektor az $\boldsymbol{A}$ mátrix oszlopaival lineárisan összefüggő. Mivel ez összesen $r+1$ darab lineárisan összefüggő vektor, létezik olyan r-ed rendű $\boldsymbol{y}$ vektor és η skalár, hogy $\boldsymbol{A}\boldsymbol{y}+\boldsymbol{b}\eta=\boldsymbol{0}$. A skalár η nem lehet nulla, mert akkor $\boldsymbol{A}$ mátrix oszlopai lineárisan összefüggők lennének, ami ellentmond a hipotézisnek. Így az $\boldsymbol{y}\eta^{-1}$ vektor létezik, és ha $\boldsymbol{x}=\boldsymbol{y}\eta^{-1}$, akkor $\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}$. Azt kell belátni, hogy az $\boldsymbol{x}$ vektor egyértelmű. Tegyük fel, hogy létezik olyan $\boldsymbol{z}$ vektor, hogy $\boldsymbol{b}=\boldsymbol{A}\boldsymbol{z}$. Ha ezt kivonjuk az előző egyenlőségből, azt kapjuk, hogy $\boldsymbol{0}=\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{z})$. Mivel $\boldsymbol{A}$ mátrix oszlopai lineárisan függetlenek, ez csak akkor lehetséges, ha $\boldsymbol{x}=\boldsymbol{z}$. □

Definíció. Az $\boldsymbol{A}$ mátrix lineárisan független oszlopainak maximum darabszáma az $\boldsymbol{A}$ mátrix rangja. Jelölése $r(\boldsymbol{A})$. Az $\boldsymbol{A}$ mátrix teljesrangú, ha rangja megegyezik az oszlopainak számával.

3.3. Nemszinguláris mátrixok

Definiáljuk a nemszinguláris mátrixot. Ehhez megvizsgáljuk a mátrix oszlopainak illetve sorainak lineáris függőségét, definiáljuk a mátrix inverzét, és megvizsgáljuk a lineáris függőség és a mátrix inverze közötti kapcsolatot.

3.3. Tétel. *Nem létezik olyan $\boldsymbol{X}$ mátrix, hogy $\boldsymbol{AX} = \boldsymbol{I}$, ha $\boldsymbol{A}$ mátrix sorai lineárisan összefüggők. [3, 30. oldal, Lemma 2.3]*

Bizonyítás. Tegyük fel az ellenkezőjét. Mivel $\boldsymbol{A}$ mátrix lineárisan összefüggő, létezik olyan $\boldsymbol{y} \neq \boldsymbol{0}$ vektor, hogy $\boldsymbol{y}^T\boldsymbol{A} = \boldsymbol{0}^T$. Így $\boldsymbol{y}^T\boldsymbol{AX} = \boldsymbol{0}^T$, de mivel $\boldsymbol{AX} = \boldsymbol{I}$, ez csak akkor teljesül, ha $\boldsymbol{y} = \boldsymbol{0}$. Így ellentmondáshoz jutottunk. □

Definíció. Legyen $\boldsymbol{A}$ négyzetes mátrix. Ha létezik olyan $\boldsymbol{X}$ mátrix, hogy

$$\boldsymbol{AX} = \boldsymbol{I}, \tag{3.8}$$

akkor $\boldsymbol{X}$ az $\boldsymbol{A}$ mátrix jobboldali inverze. Ha létezik olyan $\boldsymbol{Y}$ mátrix, hogy

$$\boldsymbol{YA} = \boldsymbol{I}, \tag{3.9}$$

akkor $\boldsymbol{Y}$ az $\boldsymbol{A}$ mátrix baloldali inverze.

Definíció. Legyen $\boldsymbol{A}$ négyzetes mátrix. Ha létezik olyan $\boldsymbol{X}$ mátrix, hogy

$$\boldsymbol{AX} = \boldsymbol{XA} = \boldsymbol{I}, \tag{3.10}$$

akkor $\boldsymbol{X}$ az $\boldsymbol{A}$ mátrix inverze. Az $\boldsymbol{A}$ mátrix inverzének jelölése $\boldsymbol{A}^{-1}$. Egy mátrix invertálható, ha létezik inverze.

3.4. Tétel. *Legyen $\boldsymbol{A}$ n-ed rendű négyzetes mátrix. Ha az $\boldsymbol{A}$ mátrix oszlopai lineárisan függetlenek, létezik olyan $\boldsymbol{X}$ mátrix, hogy $\boldsymbol{AX} = \boldsymbol{I}$. [3, 30. oldal, Theorem 2.3]*

Bizonyítás. Legyen $\boldsymbol{A}$ $n\times n$-es négyzetes mátrix, és legyenek az oszlopai lineárisan függetlenek. Legyenek $\boldsymbol{A}_i$ $(i = 1,2,...,n-1)$ $(n-1)\times n$-es mátrixok. Az $\boldsymbol{A}_i$ mátrix mindig legyen az $\boldsymbol{A}$ mátrix i-edik sorának elhagyásával kapott mátrix. Azaz $\boldsymbol{A}_i$ j-edik sora az $\boldsymbol{A}$ mátrix j-edik sora, ha $1 \leq j \leq i-1$, de a $(j+1)$-edik sora, ha $i \leq j \leq n-1$. A 3.1-es tételből következik, hogy mindig létezik olyan $\boldsymbol{x}_i \neq \boldsymbol{0}$, hogy $\boldsymbol{A}_i\boldsymbol{x}_i = \boldsymbol{0}$. Mivel létezik olyan $\boldsymbol{x}_i$, hogy $\boldsymbol{A}_i\boldsymbol{x}_i = \boldsymbol{0}$, de $\boldsymbol{Ax}_i \neq \boldsymbol{0}$, így az $\boldsymbol{Ax}_i$ vektornak egyedül az i-edik eleme nem nulla. Minden i-re választhatjuk $\boldsymbol{x}_i$-t úgy, hogy $\boldsymbol{Ax}_i = \boldsymbol{e}_i$ legyen, ahol $\boldsymbol{e}_i$ az n-ed rendű egységmátrix i-edik oszlopa. Legyen $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n]$, így $\boldsymbol{AX} = \boldsymbol{I}$. □

Megmutattuk, hogy ha az $\boldsymbol{A}$ négyzetes mátrixnak az oszlopai lineárisan függetlenek, akkor létezik jobboldali inverze. A 3.3-as tételből következik, hogy ekkor az $\boldsymbol{A}$ mátrix sorai is lineárisan függetlenek. Tehát a négyzetes mátrixok sorainak és oszlopainak lineáris függetlensége egyenértékű.

3.5. Tétel. *Ha egy négyzetes mátrix oszlopai lineárisan függetlenek, akkor jobboldali inverze egyértelmű. [3, 31. oldal, Theorem 2.4]*

Bizonyítás. Legyen $\boldsymbol{A}$ négyzetes mátrix, és legyenek az oszlopai lineárisan függetlenek. Tegyük fel, hogy az $\boldsymbol{A}$ mátrix jobboldali inverze nem egyértelmű, azaz léteznek olyan $\boldsymbol{X} \neq \boldsymbol{Y}$ mátrixok, hogy $\boldsymbol{AX} = \boldsymbol{I}$ és $\boldsymbol{AY} = \boldsymbol{I}$. Vonjuk ki a két egyenlőséget egymásból, így kapjuk, hogy $\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{Y}) = \boldsymbol{0}$. Ez azonban az $\boldsymbol{A}$ mátrix oszlopainak lineáris függetlensége miatt csak akkor lehetséges, ha $\boldsymbol{X} = \boldsymbol{Y}$. □

3.6. Tétel. *Ha egy négyzetes mátrixnak létezik egyértelmű jobboldali inverze, akkor az megegyezik a baloldali inverzével. [3, 31. oldal, Theorem 2.5]*

Bizonyítás. Legyen $\boldsymbol{A}$ négyzetes mátrix, és legyenek az oszlopai lineárisan függetlenek. Legyen $\boldsymbol{AX} = \boldsymbol{I}$. Az egyenlőséget jobbról $\boldsymbol{A}$ mátrixszal szorozva kapjuk, hogy $\boldsymbol{AXA} = \boldsymbol{A}$, átrendezve $\boldsymbol{A}(\boldsymbol{XA} - \boldsymbol{I}) = \boldsymbol{0}$. Ebből az $\boldsymbol{A}$ mátrix oszlopainak lineáris függetlensége miatt következik, hogy $\boldsymbol{XA} = \boldsymbol{I}$. □

Definíció. Legyen $\boldsymbol{A}$ négyzetes mátrix. Ha a következő ekvivalens állítások teljesülnek, akkor az $\boldsymbol{A}$ mátrix nemszinguláris, egyébként szinguláris.

1. Az $\boldsymbol{A}$ mátrix oszlopai lineárisan függetlenek.
2. Az $\boldsymbol{A}$ mátrix sorai lineárisan függetlenek.
3. Az $\boldsymbol{A}$ mátrix invertálható.

3.7. Tétel. *Legyen $\boldsymbol{A}$ négyzetes mátrix, és legyen*

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{bmatrix},$$

ahol $\boldsymbol{A}_{11}$ és $\boldsymbol{A}_{22}$ minormátrixok is négyzetes mátrixok. Ekkor az $\boldsymbol{A}$ mátrix akkor és csak akkor nemszinguláris, ha $\boldsymbol{A}_{11}$ és $\boldsymbol{A}_{22}$ minormátrixok nemszingulárisak. [3, 33. oldal, Lemma 2.4]

Bizonyítás. Legyenek $\boldsymbol{A}_{11}$ és $\boldsymbol{A}_{22}$ minormátrixok nemszingulárisak. Ekkor

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{A}_{22}^{-1} \end{bmatrix},$$

tehát az $\boldsymbol{A}$ mátrix nemszinguláris.

Legyen $\boldsymbol{A_{11}}$ minormátrix szinguláris. Ekkor mindig létezik olyan $\boldsymbol{x} \neq \boldsymbol{0}$ vektor, hogy $\boldsymbol{A}_{11}\boldsymbol{x} = \boldsymbol{0}$. Ekkor az $\boldsymbol{A}$ mátrix szingularitása csak $\boldsymbol{A}_{11}$ minormátrixtól függ, mert

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}.$$

Ezért, ha $\boldsymbol{A_{11}}$ minormátrix szinguláris, akkor az $\boldsymbol{A}$ mátrix szinguláris. Hasonlóan, ha $\boldsymbol{A}_{22}$ minormátrix szinguláris, mindig létezik olyan $\boldsymbol{y} \neq \boldsymbol{0}$ vektor, hogy $\boldsymbol{y}^T\boldsymbol{A}_{22} = \\ = \boldsymbol{0}^T$. □

Most definiálunk és megvizsgálunk néhány elméleti vagy gyakorlati szempontból fontos nemszinguláris mátrixot.

Definíció. Az ortogonális mátrix olyan valós mátrix, melynek az inverze a transzponáltja.

Definíció. Az $\boldsymbol{U} = [u_{ij}]$ négyzetes mátrix felső háromszögmátrix, ha $u_{ij} = 0$ minden $i > j$-re.

Definíció. Az $\boldsymbol{L} = [l_{ij}]$ négyzetes mátrix alsó háromszögmátrix, ha $l_{ij} = 0$ minden $i < j$-re.

Definíció. Az alsó háromszögmátrix alsó egység háromszögmátrix, ha minden diagonális eleme egy.

Definíció. A felső háromszögmátrix felső egység háromszögmátrix, ha minden diagonális eleme egy.

3.8. Tétel. *Az $\boldsymbol{U} = [u_{ij}]$ felső háromszögmátrix akkor és csak akkor nemszinguláris, ha $u_{ii} \neq 0$ minden i-re. [3, 34. oldal, Theorem 2.6]*

Bizonyítás. Legyen $\boldsymbol{U}$ n-ed rendű felső háromszögmátrix. Legyen $\boldsymbol{U}_i$ a bal felső i-ed rendű sarokminormátrixa $\boldsymbol{U}$-nak., azaz $\boldsymbol{U} = \boldsymbol{U}_n$. Legyen

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_{i-1} & \boldsymbol{v}_i \\ \boldsymbol{0} & u_{ii} \end{bmatrix}, \quad 2 \leq i \leq n$$

és $\boldsymbol{v}_i^T = [u_{1i}, u_{2i}, ..., u_{(i-1)i}]$. A 3.7-es tételből következik, hogy $\boldsymbol{U}$ akkor és csak akkor nemszinguláris, ha minden $\boldsymbol{U}_i$ sarokminormátrix nemszinguláris. Ha $u_{ii} \neq 0$ bármely $1 \leq i \leq n$ esetén, akkor $\boldsymbol{U}_1, \boldsymbol{U}_2, ..., \boldsymbol{U}_n$ sarokminormátrixok nemszingulárisak, azaz $\boldsymbol{U}$ nemszinguláris. □

3.9. Tétel. *Négyzetes mátrixok szorzata akkor és csak akkor nemszinguláris, ha a szorzat minden tényezője nemszinguláris.*

Bizonyítás. Legyenek $\boldsymbol{A}$, $\boldsymbol{B}$ és $\boldsymbol{C}$ négyzetes mátrixok. Legyen $\boldsymbol{C} = \boldsymbol{AB}$. Tegyük fel, hogy $\boldsymbol{B}$ és $\boldsymbol{C}$ mátrixok nemszingulárisak, de $\boldsymbol{A}$ mátrix szinguláris. Mivel $\boldsymbol{C}$ nemszinguláris, létezik inverze, és $\boldsymbol{CC}^{-1} = \boldsymbol{I}$. Ebből következik, hogy $(\boldsymbol{AB})\boldsymbol{C}^{-1} = \boldsymbol{I}$. De ez csak akkor lehetséges, ha $\boldsymbol{C}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$, amivel ellentmondáshoz jutottunk. □

Definíció. Az $\boldsymbol{A}$ valós mátrix pozitív definit, ha $\boldsymbol{x}^T\boldsymbol{Ax} > 0$, minden $\boldsymbol{x} \neq \boldsymbol{0}$ vektorra. Jelölése $\boldsymbol{A} > 0$.

Definíció. Az $\boldsymbol{A}$ valós mátrix pozitív szemidefinit, ha $\boldsymbol{x}^T\boldsymbol{Ax} \geq 0$, minden $\boldsymbol{x} \neq \boldsymbol{0}$ vektorra. Jelölése $\boldsymbol{A} \geq 0$.

Definíció. Az $\boldsymbol{A}$ valós mátrix negatív definit, ha $\boldsymbol{x}^T\boldsymbol{Ax} < 0$, minden $\boldsymbol{x} \neq \boldsymbol{0}$ vektorra. Jelölése $\boldsymbol{A} < 0$.

Definíció. Az $\boldsymbol{A}$ valós mátrix negatív szemidefinit, ha $\boldsymbol{x}^T\boldsymbol{Ax} \leq 0$, minden $\boldsymbol{x} \neq \boldsymbol{0}$ vektorra. Jelölése $\boldsymbol{A} \leq 0$.

Definíció. Az $\boldsymbol{A}$ valós mátrix definit, ha $\boldsymbol{x}^T\boldsymbol{Ax} \neq 0$, minden $\boldsymbol{x} \neq \boldsymbol{0}$ vektorra.

Definíció. Az $\boldsymbol{A}$ valós mátrix indefinit, ha léteznek olyan $\boldsymbol{x} \neq \boldsymbol{0}$ és $\boldsymbol{y} \neq 0$ vektorok, amelyekre $\boldsymbol{x}^T\boldsymbol{Ax} < 0 < \boldsymbol{y}^T\boldsymbol{Ay}$.

3.10. Tétel. *Legyen $\boldsymbol{A}$ $m\times n$-es valós mátrix lineárisan független oszlopokkal. Ekkor $\boldsymbol{A}^T\boldsymbol{A}$ pozitív definit. [3, 34. oldal, Lemma 2.5]*

Bizonyítás. Legyen $\boldsymbol{A}$ $m\times n$-es valós mátrix lineárisan független oszlopokkal. Az $\boldsymbol{A}^T\boldsymbol{A}$ mátrix szimmetrikus. Legyen $\boldsymbol{y} = \boldsymbol{Ax}$, így $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{y}^T\boldsymbol{y} > 0$, akkor és csak akkor, ha $\boldsymbol{y} \neq \boldsymbol{0}$. Mivel az $\boldsymbol{A}$ mátrix oszlopai lineárisan függetlenek, $\boldsymbol{y}$ akkor és csak akkor nulla, ha $\boldsymbol{x}$ is nulla. Az $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax}$ akkor és csak akkor nulla, ha $\boldsymbol{x} = \boldsymbol{0}$. □

3.4. Vektornormák és mátrixnormák

Gyakran akarunk vektorokat vagy mátrixokat a nagyságuk alapján összehasonlítani. Például ha egy iteratív eljárással közelítünk egy vektorhoz, a hibavektor (azaz a vektor és a közelítő vektor közötti különbség) nagysága jó ha gyorsan csökken. A norma a vektorokhoz és a mátrixokhoz egy skalárt rendel.

Definíció. Az $\|\boldsymbol{x}\|$ skalár az $\boldsymbol{x}$ vektor normája, ha kielégíti a következő három feltételt.

1. $\|\boldsymbol{x}\| = 0$, ha $\boldsymbol{x} = \boldsymbol{0}$, egyébként $\|\boldsymbol{x}\| > 0$.
2. $\|\boldsymbol{x}\theta\| = \|\boldsymbol{x}\|\,|\theta|$, ahol θ skalár.

3. $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Definíció. Az $\boldsymbol{x}$ vektor l_1, l_2, l_∞ normáinak definíciója:

1. $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$ az l_1 norma,
2. $\|\boldsymbol{x}\|_2 = (\sum_i |x_i^2|)^{\frac{1}{2}}$ az l_2 (euklideszi) norma,
3. $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$ az l_∞ (maximum) norma.

3.11. Tétel. *(Cauchy-egyenlőtlenség) Legyen $\boldsymbol{x}$ és $\boldsymbol{y}$ n-ed rendű nem nulla vektor. Az*

$$|\boldsymbol{x}^T\boldsymbol{y}| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \tag{3.11}$$

egyenlőtlenség akkor és csak akkor igaz, ha $\boldsymbol{y}$ vektor skalárszorosa $\boldsymbol{x}$ vektornak. [3, 42. oldal, Cauchy's Inequality]

A Cauchy-egyenlőtlenség komplex vektorokra is igaz, ha a transzponáltat Hermite-féle transzponáltra cseréjük. A tételt itt nem bizonyítjuk.

Definíció. Az $\|\boldsymbol{A}\|$ skalár az $\boldsymbol{A}$ mátrix normája, ha kielégíti a következő négy feltételt.

1. $\|\boldsymbol{A}\| = 0$, ha $\boldsymbol{A} = \boldsymbol{0}$, egyébként $\|\boldsymbol{A}\| > 0$.
2. $\|\boldsymbol{A}\theta\| = \|\boldsymbol{A}\|\,|\theta|$, ahol θ skalár.
3. $\|\boldsymbol{A}+\boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$.
4. $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$.

Definíció. Az $\boldsymbol{A}$ mátrix $\|\boldsymbol{A}\|_p$ indukált normájának definíciója

$$\boldsymbol{A} = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}, \tag{3.12}$$

ahol $p = 1{,}2$ vagy ∞.

Az indukált mátrixnormákra mindig igaz, hogy $\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{x}\|$. Ez a gyakorlatban sokszor hasznos. Az l_1 indukált mátrixnormát explicit az

$$\|\boldsymbol{A}\|_1 = \max_j \sum_i |a_{ij}| \tag{3.13}$$

képlettel számolhatjuk. Az l_∞ indukált mátrixnormát az

$$\|\boldsymbol{A}\|_\infty = \max_i \sum_j |a_{ij}| \tag{3.14}$$

képlettel számolhatjuk. Az l_2 által indukált mátrixnormát spektrális normának hívjuk. A spektrális norma elméleti fontossága mellett nagy gyakorlati hátránya, hogy nincs egyszerű explicit képlet a kiszámolására.

Definíció. Az $\boldsymbol{A}$ mátrix Frobenius normájának definíciója

$$\|\boldsymbol{A}\|_F = (\sum_{i,j} |a_{ij}|^2)^{\frac{1}{2}}. \tag{3.15}$$

A Frobenius norma az euklideszi vektornorma mátrix megfelelője, de nem az euklideszi vektornorma által indukált mátrixnorma. Nem vektornorma által indukált mátrixnorma, ezért a gyakorlatban sokszor indokolatlanul pontatlan eredményekhez vezet.

Definíció. A nemszinguláris $\boldsymbol{A}$ mátrix kondíciószámának jelölése $k(\boldsymbol{A})$, és definíciója

$$k(\boldsymbol{A}) = \|\boldsymbol{A}\| \left\|\boldsymbol{A}^{-1}\right\| , \tag{3.16}$$

ahol tetszőleges norma választható.

A kondíciószám mérete jellemzi a mátrix szingularitásának mértékét. Ha a kondíciószám kicsi, szokás a mátrixot jól kondicionáltnak hívni, ha a kondíciószám nagy, szokás a mátrixot rosszul kondicionáltnak, vagy közel szinguláris nak hívni [3]. Ezt bővebben lásd a 8.2. fejezetben.

4. Lineáris egyenletrendszerek megoldása, osztályozása

A megoldása az

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{4.1}$$

alakú egyenletrendszernek, ahol $\boldsymbol{A}$ n-ed rendű nemszinguláris mátrix, $\boldsymbol{x}$ ismeretlen n-ed rendű vektor, $\boldsymbol{b}$ tetszőleges n-ed rendű vektor, az egyik leggyakoribb feladat. Az $\boldsymbol{A}$ mátrix az egyenletrendszer együtthatómátrixa. A (4.1) egyenletrendszernek akkor és csak akkor létezik megoldása, ha a $\boldsymbol{b}$ vektor lineárisan kifejezhető az $\boldsymbol{A}$ mátrix oszlopvektoraiból. Vagy másként fogalmazva, a $\boldsymbol{b}$ vektor és az $\boldsymbol{A}$ mátrix oszlopvektorai nem lineárisan függetlenek. Ekkor az egyenletrendszer megoldása

$$\boldsymbol{A}^{-1}\boldsymbol{b}\,. \tag{4.2}$$

Az $\boldsymbol{A}$ mátrix inverzének számolása a gyakorlatban nem célszerű, mert túl műveletigényes. A megoldási módszerek két fő osztálya az iteratív eljárások és a direkt

eljárások. Egy megoldási módszer hatékonysága két fő szempont alapján bírálható el:

1. A megoldási módszer mennyire gyors, azaz mennyire műveletigényes?
2. Mennyire pontos a kiszámított megoldás? [14]

A gyakorlatban előforduló együtthatómátrixok általában vagy kitöltöttek és alacsony rendszámúak (pl. a rendszám kisebb mint 30 [14]), vagy ritkák és nagy rendszámúak. A direkt eljárások általában előnyösebbek a kis rendszámú kitöltött mátrixoknál, míg az iteratív eljárások általában előnyösebbek a nagy rendszámú ritka mátrixoknál. [14]

Az egyenletrendszer megoldásánál számolnunk kell különböző hibaforrásokkal. Hibaforrás lehet az együtthatómátrix vagy a jobboldali $\boldsymbol{b}$ vektor elemeinek mérési hibája, a számítások közben keletkező kerekítési hibák, és a megoldási módszer képlethibája. [14]

4.1. Direkt eljárások

A direkt eljárások elvileg mentesek a képlethibától, és a megoldást véges sok lépésben állítják elő. Elméletben a megoldás pontos, a gyakorlatban azonban ez mégsem teljesül, a lépések közben keletkező kerekítési hibák miatt. A direkt eljárások alapja az (4.1) egyenletrendszer sorozatos transzformációja, amíg a megoldás könnyen számolhatóvá válik.

4.1.1. Az LU-felbontás

Definíció. Az $\boldsymbol{A}$ mátrix LU-felbontásán az

$$\boldsymbol{A} = \boldsymbol{LU} \tag{4.3}$$

felbontást értjük, ahol $\boldsymbol{L}$ alsó háromszögmátrix és $\boldsymbol{U}$ felső háromszögmátrix.

Az LU-felbontással az (4.1) egyenletrendszer megoldása az

$$\boldsymbol{Ly} = \boldsymbol{b} \tag{4.4}$$

és

$$\boldsymbol{Ux} = \boldsymbol{y} \tag{4.5}$$

háromszög alakú egyenletrendszerek megoldására egyszerűsíthető, ahol $\boldsymbol{y}$ vektor a felbontás egy köztes eredménye. Először a (4.5) egyenletrendszert oldjuk meg. Legyen

$\boldsymbol{U} = [u_{ij}]$, $\boldsymbol{x} = [x_i]$ és $\boldsymbol{y} = [y_i]$ $(i, j = 1,2, ..., n)$ ekkor

$$x_i = (y_i - \sum_{j=i+1}^{n} u_{ij}x_j)/u_{ii}\,, \tag{4.6}$$

ahol $u_{ii} \neq 0$. Ahhoz, hogy a (4.5) egyenletrendszernek tetszőleges $\boldsymbol{y}$ vektorra legyen megoldása, feltételezzük, hogy $\boldsymbol{U}$ mátrix nemszinguláris. Így a 3.8-es tételből következik, hogy $u_{ii} \neq 0$. A (4.4) egyenletrendszer $\boldsymbol{y}$ vektor ismeretében hasonlóan oldható meg. [3]

4.1. Tétel. *Az $\boldsymbol{A}$ nemszinguláris mátrixnak akkor és csak akkor létezik LU-felbontása, ha az $\boldsymbol{A}$ mátrix minden bal felső sarokminormátrixa nemszinguláris. [3, 56. oldal, Theorem 4.1]*

Bizonyítás. Legyen $\boldsymbol{A}$ mátrix nemszinguláris. Megmutatjuk, hogy ha az $\boldsymbol{A} = \boldsymbol{LU}$ felbontás létezik, akkor az $\boldsymbol{A}$ mátrix bal felső sarokminormátrixai szükségképpen nemszingulárisak. Particionáljuk $\boldsymbol{A}$ mátrixot úgy, hogy

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix},$$

ahol $\boldsymbol{A}_{11}$ egy tetszőleges rendű bal felső sarokminormátrixa az $\boldsymbol{A}$ mátrixnak. Particionáljuk az $\boldsymbol{L}$ és $\boldsymbol{U}$ mátrixokat hasonlóan. A hipotézis szerint $\boldsymbol{A} = \boldsymbol{LU}$, azaz

$$\begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{0} & \boldsymbol{U}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$$

és

$$\boldsymbol{L}_{11}\boldsymbol{U}_{11} = \boldsymbol{A}_{11}\,.$$

Az $\boldsymbol{A}$ mátrix nemszinguláris, így a 3.9-es tétel miatt $\boldsymbol{L}$ és $\boldsymbol{U}$ mátrixok is nemszingulárisak. A 3.7-es tétel miatt $\boldsymbol{L}_{11}$ és $\boldsymbol{U}_{11}$ szintén nemszingulárisak, így $\boldsymbol{A}_{11}$ nemszinguláris. Mivel $\boldsymbol{A}_{11}$ egy tetszőleges rendű bal felső sarokminormátrixa az $\boldsymbol{A}$ mátrixnak, így az $\boldsymbol{A}$ mátrix szükségképpen nemszinguláris.

Legyen

$$\boldsymbol{A}_r = \boldsymbol{L}_r\boldsymbol{U}_r\,, \tag{4.7}$$

ahol $\boldsymbol{A}_r, \boldsymbol{L_r}$ és $\boldsymbol{U}_r$ az $\boldsymbol{A}, \boldsymbol{L}$ és $\boldsymbol{U}$ mátrixok r-ed rendű bal felső sarokminormátrixai. Megmutatjuk, hogy ha $\boldsymbol{A}_r$ nemszinguláris, akkor $\boldsymbol{A}_{r+1}$ is szükségképpen nemszinguláris. Indukcióval bizonyítunk. Particionáljuk $\boldsymbol{A}_{r+1}$-t úgy, hogy

$$\boldsymbol{A}_{r+1} = \begin{bmatrix} \boldsymbol{A}_r & \boldsymbol{b}_r \\ \boldsymbol{c}_r^T & \delta_r \end{bmatrix}, \tag{4.8}$$

ahol $\boldsymbol{b}_r$ és $\boldsymbol{c}_r$ r-ed rendű vektorok és δ_r skalár. Particionáljuk $\boldsymbol{L}_{r+1}$-et és $\boldsymbol{U}_{r+1}$-et úgy, hogy

$$\boldsymbol{L}_{r+1} = \begin{bmatrix} \boldsymbol{L}_r & \boldsymbol{0} \\ \boldsymbol{p}_r & \lambda_r \end{bmatrix} \tag{4.9}$$

és

$$\boldsymbol{U}_{r+1} = \begin{bmatrix} \boldsymbol{U}_r & \boldsymbol{v}_r \\ \boldsymbol{0}^T & \mu_r \end{bmatrix}. \tag{4.10}$$

Ekkor található olyan $\boldsymbol{p}_r, \lambda_r, \boldsymbol{v}_r$ és μ_r, hogy

$$\boldsymbol{L}_{r+1}\boldsymbol{U}_{r+1} = \boldsymbol{A}_{r+1}\,. \tag{4.11}$$

A particionált (4.8), (4.9) és (4.10) mátrixokkal felírva a (4.11) egyenlőséget

$$\begin{bmatrix} \boldsymbol{L}_r & \boldsymbol{0} \\ \boldsymbol{p}_r & \lambda_r \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_r & \boldsymbol{v}_r \\ \boldsymbol{0}^T & \mu_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_r & \boldsymbol{b}_r \\ \boldsymbol{c}_r^T & \delta_r \end{bmatrix}.$$

A szorzás elvégzése után a (4.7) egyenlőséget és az

$$\boldsymbol{L}_r\boldsymbol{v}_r = \boldsymbol{b}_r\,, \tag{4.12}$$

$$\boldsymbol{p}_r^T\boldsymbol{U}_r = \boldsymbol{c}_r^T \tag{4.13}$$

és

$$\boldsymbol{p}_r^T\boldsymbol{v}_r + \lambda_r\mu_r = \delta_r\,, \tag{4.14}$$

egyenlőségeket kapjuk. Mivel a hipotézis szerint $\boldsymbol{A}_r$ nemszinguláris, a (4.7) egyenlőségből következik, hogy $\boldsymbol{L}_r$ és $\boldsymbol{U}_r$ szintén nemszingulárisak. Ezért léteznek olyan egyértelmű $\boldsymbol{p}_r$ és $\boldsymbol{v}_r$ vektorok, amelyek kielégítik a (4.12) és (4.13) egyenlőségeket. Mivel léteznek olyan λ_r és μ_r skalárok, amik kielégítik a (4.14) egyenlőséget, létezik olyan $\boldsymbol{L}_{r+1}$ és $\boldsymbol{U}_{r+1}$, amik kielégítik a (4.11) egyenlőséget. Tehát, ha az $\boldsymbol{A}$ mátrix bal felső r-ed rendű sarokminormátrixa nemszinguláris, és az $\boldsymbol{A}$ mátrixnak létezik LU-felbontása, akkor az $\boldsymbol{A}$ mátrix bal felső $r+1$-ed rendű sarokminormátrixa is nemszinguláris. □

Következmény. *Ha $\boldsymbol{L}$ alsó egység háromszögmátrix, az LU-felbontás egyértelmű. [3, 57. oldal, Corollary]*

Bizonyítás. Ha $\boldsymbol{L}$ alsó egység háromszögmátrix, akkor a (4.9) felbontásban $\lambda_r = 1$ minden r-re. Ebből következik, hogy (4.10) felbontásban μ_r egyértelmű minden r-re. Mivel $\boldsymbol{p}_r$ és $\boldsymbol{v}_r$ egyértelmű minden r-re, következik az eredmény. □

4.1.2. A Gauss-féle elimináció

Az egyik legrégebbi és legjobb módszer a (4.1) egyenletrendszer megoldására a Gauss-féle elimináció. Az első, második, stb. egyenlőség megfelelő skalárszorosát kivonjuk a többi egyenlőségből, úgy, hogy ezzel az ismeretlen változókat eliminájuk. Ezt addig folytatjuk, amíg az utolsó egyenlőségben már csak egy ismeretlen változó marad. Ekkor a kapott egyenlőségek felső háromszög alakúak. A módszert egy példán mutatjuk be.

Példa. Oldjuk meg a

$$\begin{array}{rrrrcr} 2x_1 & +x_2 & +3x_3 & -x_4 & = & 3 \\ -4x_1 & -3x_2 & -4x_3 & +5x_4 & = & 2 \\ 6x_1 & +4x_2 & +4x_3 & -5x_4 & = & -1 \\ -4x_1 & -3x_2 & +2x_3 & +4x_4 & = & 7 \end{array}$$

egyenletrendszert! Vonjuk ki az 1. egyenlőség $(-2)-$szeresét a második és negyedik egyenlőségből, és a 3-szorosát a harmadik egyenlőségből. Így ezekből az egyenlőségekből az x_1 ismeretlent eliminátuk, és a

$$\begin{array}{rrrrcr} 2x_1 & +x_2 & +3x_3 & -x_4 & = & 3 \\ & -x_2 & +2x_3 & +3x_4 & = & 8 \\ & x_2 & -5x_3 & -2x_4 & = & -10 \\ & -3x & +8x_3 & +2x_4 & = & 13 \end{array}$$

új egyenletrendszert kaptuk. Az új egyenletrendszerben az x_2 ismeretlen eliminálásához vonjuk ki a 2. egyenlőség (-1)-szeresét a harmadik egyenlőségből, és az 1-szeresét a negyedik egyenlőségből. Így ezekből az egyenlőségekből az x_2 ismeretlent eliminátuk, és a

$$\begin{array}{rrrrcr} 2x_1 & +x_2 & +3x_3 & -x_4 & = & 3 \\ & -x_2 & +2x_3 & +3x_4 & = & 8 \\ & & -3x_3 & x_4 & = & -2 \\ & & +6x_3 & x_4 & = & 5 \end{array}$$

új egyenletrendszert kaptuk. Végül vonjuk ki a 3. egyenlőség (-2)-szeresét a negyedik egyenlőségből. Így a negyedik egyenlőségből x_2-t eliminátuk, és a

$$\begin{array}{rrrrcr} 2x_1 & +x_2 & +3x_3 & -x_4 & = & 3 \\ & -x_2 & +2x_3 & +3x_4 & = & 8 \\ & & -3x_3 & x_4 & = & -2 \\ & & & x_4 & = & 1 \end{array}$$

új egyenletrendszert kaptuk. Visszahelyettesítve az $x_4 = 1$-et kapjuk, hogy $x_3 = 1$, $x_2 = -3$ és $x_1 = 2$. [3, 60. oldal, Example 4.3]

Legyen $\boldsymbol{A}^{(r-1)} = [a_{ij}^{(r-1)}]$, $1 \leq r \leq n$ a (4.1) egyenletrendszer Gauss-féle eliminációval való megoldásakor az $(r-1)$-edik eliminációs lépés után kapott együtthatómátrix. Legyen $\boldsymbol{M}_r = [m_{ij}]$ és

$$\boldsymbol{A}^{(r)} = \boldsymbol{M}_r \boldsymbol{A}^{(r-1)}. \tag{4.15}$$

A (4.15) egyenlőségben az $\boldsymbol{M}_r$ mátrixot mindig úgy választjuk, hogy a szorzás eredményeként az $\boldsymbol{A}^{(r)}$ együtthatómátrix r-edik oszlopában az r-edik sor alatt az együtthatókra nullákat kapjunk. Ehhez a (4.15) egyenlőségben

$$\boldsymbol{M}_r = \boldsymbol{I} - \boldsymbol{m}_r \boldsymbol{e}_r^T, \tag{4.16}$$

ahol

$$\boldsymbol{m}_r = [0,0,...,0, m_{r+1,r}, m_{r+2,r}, .., m_{nr}]^T$$

és

$$m_{ir} = \frac{a_{ir}^{(r-1)}}{a_{rr}^{(r-1)}}, \qquad a_{rr}^{(r-1)} \neq 0, \tag{4.17}$$

az $\boldsymbol{e}_r$ vektor pedig az r-edik oszlopa az $\boldsymbol{I}$ egységmátrixnak. Ekkor az $\boldsymbol{A}^{(r)}$ mátrixot úgy számolhatjuk, hogy kivonjuk az $\boldsymbol{A}^{(r-1)}$ mátrix r-edik sorának m_{ir}-szeresét az $\boldsymbol{A}^{(r-1)}$ mátrix i-edik sorából minden i-re, ahol $i = r+1, r+2, ..., n$. [3]

Definíció. A (4.15) egyenlőségben az m_{ir} elemeket multiplikátoroknak hívjuk, az $a_{rr}^{(r-1)}$ elemeket pedig r-edik pivot, vagy r-edik főelemnek hívjuk.

A módszer az alkalmazásakor az

$$\boldsymbol{A}^{(r)}\boldsymbol{x} = \boldsymbol{b}^{(r)} \tag{4.18}$$

egyenlőségrendszereket generálja, ahol $\boldsymbol{b}^{(r)} = \boldsymbol{M}_r \boldsymbol{b}^{(r-1)}$.

$$\boldsymbol{A}^{(n-1)} = \boldsymbol{M}_{n-1}\boldsymbol{M}_{n-2}...\boldsymbol{M}_1\boldsymbol{A}, \tag{4.19}$$

ahol $\boldsymbol{A}^{(n-1)}$ felső háromszögmátrix, és $\boldsymbol{M}_r$ a (4.16) egyenlőség miatt mindig alsó egység háromszögmátrix. Alsó egység háromszögmátrix inverze alsó egység háromszögmátrix, és alsó egység háromszögmátrixok szorzata is alsó egység háromszögmátrix, így ha

$$\boldsymbol{U} = \boldsymbol{A}^{(n-1)} \tag{4.20}$$

és

$$\boldsymbol{L} = (\boldsymbol{M}_{n-1}\boldsymbol{M}_{n-2}...\boldsymbol{M}_1)^{-1}, \tag{4.21}$$

a módszer eredménye az $\boldsymbol{A}$ együtthatómátrix LU-felbontását adja. Mivel $\boldsymbol{L}$ alsó egység háromszögmátrix, az LU-felbontás egyértelmű. Ha $\boldsymbol{x}$ kielégíti a (4.18) egyenlőséget, akkor kielégíti a (4.1) egyenlőséget is. [3]

A (4.17) egyenlőségből látszik, hogy a pivot elemek nem lehetnek nullák. Ha bármelyik pivot elem nulla, akkor az $\boldsymbol{A}$ mátrix bal felső sarokminormátrixa szinguláris. Ebből következik, hogy az $\boldsymbol{A}$ mátrixnak nem létezik LU-felbontása (4.1-es tétel). [3]

A nulla értékű r-edik pivot elem kiküszöbölésének egy módja, ha az $\boldsymbol{A}^{(r-1)}$ mátrix sorait felcseréljük, úgy hogy az r-edik pivot elem helyén ne nullát kapjunk. Ezt úgy érhetjük el, ha az $\boldsymbol{A}^{(r-1)}$ mátrix r-edik oszlopában az $a_{rr}^{(r-1)}$ elemtől lefelé megkeressük az első nem nulla elemet, legyen ez pl. $a_{ir}^{(r-1)}$, majd felcseréljük az $\boldsymbol{A}^{(r-1)}$ mátrix r-edik sorát az i-edik sorával. Ezt megtehetjük, mert $i > r$. [3]

4.2. Iteratív eljárások

Az iteratív megoldási módszereknél a (4.1) egyenletrendszer $\boldsymbol{A}$ mátrixa lényegében változatlan marad és közelítő $\boldsymbol{x}_i$ $(i = 1,2,...)$ megoldásokat generálunk. Azt reméljük, hogy az $\{\boldsymbol{x}_i\}$ sorozat a megoldáshoz konvergál. A közelítő megoldások generálásának leállítását, azaz az iterációk leállítását, kilépési feltételhez, feltételekhez kötjük.

Definíció. Az $\boldsymbol{A}$ mátrixot $\boldsymbol{P} - \boldsymbol{Q}$ alakban kifejezve, az $\boldsymbol{A}$ mátrix egy felbontásának nevezzük.

Legyen az $\boldsymbol{A} = \boldsymbol{P} - \boldsymbol{Q}$ olyan felbontás, ahol $\boldsymbol{P}$ nemszinguláris, és a $\boldsymbol{Py} = \boldsymbol{c}$ egyenletrendszer egyszerűen megoldható. Írjuk a (4.1) egyenletrendszert

$$\boldsymbol{Px} = \boldsymbol{b} + \boldsymbol{Qx} \tag{4.22}$$

alakban. Az $\boldsymbol{x}$ vektort úgy próbáljuk meghatározni, hogy egy tetszőleges $\boldsymbol{x}_0$ becslésből kiindulva egy $\{\boldsymbol{x}_i\}$ sorozatot generálunk, a

$$\boldsymbol{Px}_{i+1} = \boldsymbol{b} + \boldsymbol{Qx}_i \tag{4.23}$$

egyenletrendszer sorozatos megoldásával. [3]

4.2. Tétel. *A (4.23) egyenlőségből képzett $\{\boldsymbol{x}_i\}$ sorozat a (4.1) egyenletrendszer megoldásához konvergál, ha $\left\|\boldsymbol{P}^{-1}\boldsymbol{Q}\right\| < 1$, tetszőleges normával. [3, 68. oldal, Theorem 4.3]*

Bizonyítás. Legyen a hibavektor $\boldsymbol{e}_i = \boldsymbol{x}_i - \boldsymbol{x}$, ahol $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, a (4.1) egyenletrendszer megoldása. A (4.22) összefüggést kivonva a (4.23) összefüggésből és balról $\boldsymbol{P}^{-1}$-el szorozva kapjuk, hogy $\boldsymbol{e}_{i+1} = \boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{e}_i$.

Legyen $\left\|\boldsymbol{P}^{-1}\boldsymbol{Q}\right\| = \alpha$, ekkor $\|\boldsymbol{e}_{i+1}\| \leq \alpha \|\boldsymbol{e}_i\|$, és ebből következik, hogy $\|\boldsymbol{e}_i\| \leq$ $\leq \alpha^i \|\boldsymbol{e}_0\|$. Tehát ha $\alpha < 1$, elég nagy i-re $\|\boldsymbol{e}_i\|$ tetszőlegesen kicsi. Azaz $\lim_{i\to\infty} \boldsymbol{x}_i =$ $= \boldsymbol{x}$. □

A fentiekben feltételeztük, hogy az $\boldsymbol{A}$ mátrix felbontása, a $\boldsymbol{P}$ és a $\boldsymbol{Q}$ mátrixok i-től függetlenek.

Definíció. Az iterációs eljárást stacionáriusnak nevezzük, ha az együtthatómátrix felbontása nem függ i-től.

Legyen $\boldsymbol{A}$ nemszinguláris mátrix és legyen $\boldsymbol{A} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$, ahol $\boldsymbol{D}$ diagonális mátrix, $\boldsymbol{L}$ alsó háromszögmátrix és $\boldsymbol{U}$ felső háromszögmátrix. Néhány példa stacionárius iterációs eljárásokra:

- Jacobi-iteráció: $\boldsymbol{P} = \boldsymbol{D}, \quad \boldsymbol{Q} = \boldsymbol{L} + \boldsymbol{U}$,
- Gauss-Seidel eljárás: $\boldsymbol{P} = \boldsymbol{D} - \boldsymbol{L}, \quad \boldsymbol{Q} = \boldsymbol{U}$,
- felső relaxációs eljárás: $\boldsymbol{P} = \omega^{-1}\boldsymbol{D} - \boldsymbol{L}, \quad \boldsymbol{Q} = \boldsymbol{U} + (\omega^{-1} - 1)\boldsymbol{D}$.

Az ω skalárral a konvergenciát gyorsítjuk. Ez általában azt jelenti, hogy $1 \leq \omega < 2$. Ha $\omega = 1$, a módszer pontosan a Gauss-Seidel eljárás. [3]

A következő fejezetekben két iterációs eljárással fogunk részletesebben foglalkozni, Broyden kutatásainak eredményeire támaszkodva. A felső relaxációs eljárással és a blokk konjugált gradiens módszerrel.

4.3. Sajátértékek és sajátvektorok

Ha az $\boldsymbol{A}$ mátrix négyzetes és nemszinguláris, akkor az $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ egyenletrendszernek egyértelmű a megoldása. Ilyen egyenletrendszerek sokszor előfordulnak, amikor valamilyen rendszer statikus viselkedését vizsgáljuk, ahol a rendszer válasza a valamilyen kísérletre a $\boldsymbol{b}$ vektor. Azonban, ha az ilyen rendszerek dinamikus viselkedését akarjuk vizsgálni, akkor meg kell határoznunk azokat a λ skalárokat, amelyekre az $(\boldsymbol{A} - \boldsymbol{I}\lambda)$ mátrix szinguláris. Ezek a λ skalárok a vizsgált rendszer belső tulajdonságaival függenek össze. [3]

Definíció. Legyen $\boldsymbol{A}$ négyzetes mátrix. A λ skalár, ami mellett az $\boldsymbol{A} - \lambda\boldsymbol{I}$ mátrix szinguláris, az $\boldsymbol{A}$ mátrix sajátértéke.

Definíció. Legyen $\boldsymbol{A}$ négyzetes mátrix. Az $\boldsymbol{x} \neq \boldsymbol{0}$ vektor, ami mellett

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}, \tag{4.24}$$

az $\boldsymbol{A}$ mátrix jobboldali sajátvektora. Legyen $\boldsymbol{B}$ négyzetes mátrix. Az $\boldsymbol{y} \neq \boldsymbol{0}$ vektor, ami mellett

$$\boldsymbol{y}^T(\boldsymbol{B} - \lambda \boldsymbol{I}) = \boldsymbol{0}^T, \tag{4.25}$$

a $\boldsymbol{B}$ mátrix baloldali sajátvektora.

A (4.24) egyenlőségből

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}. \tag{4.26}$$

Definíció. Legyen $\boldsymbol{A}$ n-ed rendű négyzetes mátrix és $\boldsymbol{x} \neq \boldsymbol{0}$ tetszőleges n-ed rendű vektor. Ekkor az

$$\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}^2\boldsymbol{x}, ...$$

sorozatot Krylov sorozatnak nevezzük.

4.3. Tétel. *Minden négyzetes mátrixnak van legalább egy sajátértéke. [3, 76. oldal, Theorem 5.1]*

Bizonyítás. Legyen $\boldsymbol{A}$ n-ed rendű négyzetes mátrix és $\boldsymbol{x} \neq \boldsymbol{0}$ tetszőleges n-ed rendű vektor. Az $\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}^2\boldsymbol{x}, ..., \boldsymbol{A}^k\boldsymbol{x}$ Krylov sorozat vektorai $k > n$ esetén lineárisan összefüggők a 3.1-es tétel miatt, de nem zárhatjuk ki, hogy a sorozat ennél kevesebb vektora is lineárisan összefüggő. Tegyük fel, hogy a sorozat első $r \leq n$ vektora lineárisan független, de az első $r + 1$ darab vektora lineárisan összefüggő. Ekkor léteznek olyan α_i, $i = 0{,}1, ..., r$, nem mind nulla, $r + 1$ darab skalárok, hogy

$$\boldsymbol{x}\alpha_0 + \boldsymbol{A}\boldsymbol{x}\alpha_1 + ... + \boldsymbol{A}^r\boldsymbol{x}\alpha_r = \boldsymbol{0}. \tag{4.27}$$

Az $\alpha_r \neq 0$, ugyanis ebből a sorozat első $r+1$ darab vektorának lineáris függetlensége következne, ami ellentmond a hipotézisnek. Ezért a (4.27) egyenlőség átirtható a

$$\left(\frac{\alpha_0}{\alpha_r}\boldsymbol{I} + \frac{\alpha_1}{\alpha_r}\boldsymbol{A} + \frac{\alpha_2}{\alpha_r}\boldsymbol{A}^2 + ... + \boldsymbol{A}^r\right)\boldsymbol{x} = \boldsymbol{0} \tag{4.28}$$

alakra. Egyszerűbben

$$p(\boldsymbol{A})\boldsymbol{x} = \boldsymbol{0},$$

ahol

$$p(\boldsymbol{A}) = \frac{\alpha_0}{\alpha_r}\boldsymbol{I} + \frac{\alpha_1}{\alpha_r}\boldsymbol{A} + ... + \boldsymbol{A}^r\,. \tag{4.29}$$

Legyen

$$p(\xi) = \frac{\alpha_0}{\alpha_r} + \frac{\alpha_1}{\alpha_r}\xi + ... + \xi^r \tag{4.30}$$

polinom, ahol ξ skalár. Az algebra alaptételéből tudjuk, hogy

$$p(\xi) = (\xi - \lambda_1)(\xi - \lambda_2)...(\xi - \lambda_r)\,, \tag{4.31}$$

ahol $\lambda_1, \lambda_1, ..., \lambda_r$ a (nem feltétlen különböző) gyökei a $p(\xi) = 0$ egyenletnek.

$$p(\boldsymbol{A}) = (\boldsymbol{A} - \lambda_1 \boldsymbol{I})(\boldsymbol{A} - \lambda_2 \boldsymbol{I})...(\boldsymbol{A} - \lambda_r \boldsymbol{I})\,. \tag{4.32}$$

Ezt onnan láthatjuk, ha a (4.31) és (4.32) egyenlőségek jobboldalait kifejtjük, és összehasonlítjuk a ξ^i és $\boldsymbol{A}^i$ együtthatókat, $0 \leq i \leq r$.

Most megmutatjuk, hogy $(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$ szinguláris. Tegyük fel az ellenkezőjét. Ekkor a (4.28) egyenlőséget balról $(\boldsymbol{A} - \lambda_1 \boldsymbol{I})^{-1}$-el megszorozva kapjuk, a (4.29) és (4.32) egyenlőséget felhasználva, hogy

$$(\boldsymbol{A} - \lambda_2 \boldsymbol{I})(\boldsymbol{A} - \lambda_3 \boldsymbol{I})...(\boldsymbol{A} - \lambda_r \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}\,, \tag{4.33}$$

ahol most $\boldsymbol{x}$ $(r-1)$-ed rendű. A (4.33) egyenlőség úgy is írható, hogy

$$(\beta_0 \boldsymbol{I} + \beta_1 \boldsymbol{A} + ... + \boldsymbol{A}^{r-1})\boldsymbol{x} = \boldsymbol{0},$$

vagy

$$\boldsymbol{x}\beta_0 + \boldsymbol{A}\boldsymbol{x}\beta_1 + ... + \boldsymbol{A}^{r-1}\boldsymbol{x} = \boldsymbol{0},$$

ahol $0 \leq i \leq r-2$, és β_i-k a megfelelő együtthatók. Ebből az következik, hogy az $\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, ...$ sorozat első r vektora lineárisan összefüggő, ami ellentmond a hipotézisnek. Ez az ellentmondás garantálja az $(\boldsymbol{A} - \lambda_1 \boldsymbol{I})$ mátrix szingularitását, és az $\boldsymbol{A}$ mátrix legalább egy sajátértékének létezését. □

4.4. Tétel. *Legyen $\boldsymbol{A}$ komplex négyzetes mátrix. Ha λ sajátértéke és $\boldsymbol{x}$ sajátvektora az $\boldsymbol{A}$ mátrixnak, akkor $\overline{\lambda}$ is sajátértéke és $\overline{\boldsymbol{x}}$ is sajátvektora $\boldsymbol{A}$ mátrixnak. [3, 77. oldal, Theorem 5.2]*

Bizonyítás. Mivel $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{z}\lambda$, a komplex konjugáltakat véve $\overline{(\boldsymbol{A}\boldsymbol{z})} = \overline{(\boldsymbol{z}\lambda)}$. $\overline{(\boldsymbol{A}\boldsymbol{z})} = \overline{\boldsymbol{A}}\overline{z}$ és $\overline{(\boldsymbol{z}\lambda)} = \overline{\lambda}\overline{z}$. □

Következmény. *Ha $\boldsymbol{A}$ valós mátrix, és ha van komplex sajátértéke és sajátvektora, ezek mindig komplex konjugált párokban jelentkeznek.*

Bizonyítás. Ha $\boldsymbol{A}$ valós mátrix, akkor $\boldsymbol{A} = \overline{\boldsymbol{A}}$. Az eredmény a tételből közvetlenül következik. □

4.5. Tétel. *Legyen az $\boldsymbol{A}$ mátrix pozitív definit. Ekkor az $\boldsymbol{A}$ mátrix minden sajátértéke pozitív.*

Bizonyítás. Tegyük fel, hogy λ sajátértéke az $\boldsymbol{A}$ mátrixnak, és $\boldsymbol{A}$ pozitív definit. Ha $\lambda = 0$, akkor létezik az $\boldsymbol{A}$ mátrixnak olyan $\boldsymbol{x}$ sajátvektora, hogy $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. De ekkor $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 0$, ami ellentmondás. Ha $\lambda < 0$, akkor létezik az $\boldsymbol{A}$ mátrixnak olyan $\boldsymbol{x}$ sajátvektora, hogy $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{A}$. De ekkor $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \lambda\,|\boldsymbol{x}|^2$, ami negatív, így ellentmondás. □

Bevezetjük a valós ortogonális mátrix komplex megfelelőjét, az unitér mátrixot.

Definíció. Az $\boldsymbol{A}$ komplex mátrix unitér mátrix, ha $\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{I}$.

A Hermite-féle mátrixot ezért a valós szimmetrikus mátrix komplex megfelelőjének tekintjük.

Definíció. Legyen $\boldsymbol{A}$ és $\boldsymbol{P}$ n-ed rendű mátrix, és legyen $\boldsymbol{P}$ nemszinguláris. A $\boldsymbol{PAP}^{-1}$ transzformációt hasonlósági transzformációnak, az $\boldsymbol{A}$ és $\boldsymbol{PAP}^{-1}$ mátrixokat hasonlónak nevezzük. Ezenfelül, ha $\boldsymbol{P}$ mátrix ortogonális (unitér), a $\boldsymbol{PAP}^{-1}$ hasonlósági transzformációt ortogonális (unitér) transzformációnak nevezzük.

4.6. Tétel. *A hasonlósági transzformáció nem változtatja meg a mátrix sajátértékeit. [3, 79. oldal, Theorem 5.3]*

Bizonyítás. Legyen $\boldsymbol{A}$ és $\boldsymbol{P}$ n-ed rendű mátrix, és legyen $\boldsymbol{P}$ nemszinguláris. Legyen $\boldsymbol{Ax} = \boldsymbol{x}\lambda$. Ekkor $\boldsymbol{PAP}^{-1}\boldsymbol{Px} = \boldsymbol{Px}\lambda$. Ezért, ha λ az $\boldsymbol{A}$ mátrixhoz tartozó sajátérték, és $\boldsymbol{x}$ a hozzátartozó sajátvektor, akkor λ a $\boldsymbol{PAP}^{-1}$ mátrixnak is sajátértéke, és $\boldsymbol{Px}$ a hozzátartozó sajátvektora. □

4.7. Tétel. *Minden k-ad rendű $\boldsymbol{A}$ mátrixhoz létezik olyan $\boldsymbol{X}$ unitér mátrix, hogy $\boldsymbol{X}^H\boldsymbol{AX}$ első oszlopa a k-ad rendű egységmátrix első oszlopának többszöröse. [3, 79. oldal, Lemma 5.2]*

Bizonyítás. Legyen $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, ahol $\boldsymbol{x}$ normalizált, azaz $\boldsymbol{x}^H\boldsymbol{x} = 1$. Legyen $\boldsymbol{X}$ olyan unitér mátrix, aminek az első oszlopa $\boldsymbol{x}$. Ekkor $\boldsymbol{x} = \boldsymbol{Xe}_1$, így $\boldsymbol{AXe}_1 = \boldsymbol{Xe}_1\lambda$. Balról szorozva $\boldsymbol{X}^H$ mátrixszal kapjuk, hogy

$$\boldsymbol{X}^H\boldsymbol{AXe}_1 = \boldsymbol{e}_1\lambda\,. \tag{4.34}$$

□

Megmutatjuk, hogy minden n-ed rendű négyzetes mátrixot hasonlósági transzformációval felső háromszögmátrixszá alakíthatunk. A felső háromszögmátrix sajátértékei pedig a főátló elemei.

4.8. Tétel. *Minden n-ed rendű $\boldsymbol{A}$ mátrixhoz létezik egy olyan $\boldsymbol{Q}$ unitér mátrix, hogy*

$$\boldsymbol{Q}^H\boldsymbol{AQ} = \boldsymbol{U}, \tag{4.35}$$

ahol $\boldsymbol{U}$ felső háromszögmátrix. [3, 80. oldal, Theorem 5.4 (Schur's theorem)]

Bizonyítás. Indukcióval bizonyítunk. Tegyük fel, hogy létezik olyan $\boldsymbol{Q}_r$ unitér mátrix, hogy

$$\boldsymbol{Q}_r^H\boldsymbol{AQ}_r = \begin{bmatrix} \boldsymbol{U}_r & \boldsymbol{B}_r \\ \boldsymbol{0} & \boldsymbol{C}_r \end{bmatrix}, \tag{4.36}$$

ahol $\boldsymbol{U}_r$ r-ed rendű felső háromszögmátrix. Megmutatjuk, hogy ekkor az egyenlőség $(r+1)$-re is igaz. Mivel $\boldsymbol{C}_r$ négyzetes, a 4.7-es tétel miatt létezik olyan $\boldsymbol{X}_r$ unitér mátrix, hogy

$$\boldsymbol{X}_r^H \boldsymbol{C}_r \boldsymbol{X}_r \boldsymbol{e}_1 = \boldsymbol{e}_1 \lambda \,, \tag{4.37}$$

ahol λ a $\boldsymbol{C}_r$ részmátrix sajátértéke. Legyen

$$\boldsymbol{P}_r = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_r \end{bmatrix}. \tag{4.38}$$

Mivel $\boldsymbol{P}_r$ unitér mátrix, a (4.36) és (4.38) egyenlőségekből következik, hogy

$$\boldsymbol{P}_r \boldsymbol{Q}_r^H \boldsymbol{A} \boldsymbol{Q}_r \boldsymbol{P}_r = \begin{bmatrix} \boldsymbol{U}_r & \boldsymbol{B}_r \boldsymbol{X}_r \\ \boldsymbol{0} & \boldsymbol{X}_r \boldsymbol{C}_r \boldsymbol{X}_r \end{bmatrix}. \tag{4.39}$$

Így a (4.37) egyenlőség miatt a (4.39) egyenlőség jobb oldala írható úgy, hogy

$$\begin{bmatrix} \boldsymbol{U}_{r+1} & \boldsymbol{B}_{r+1} \\ \boldsymbol{0} & \boldsymbol{C}_{r+1} \end{bmatrix}.$$

Legyen $\boldsymbol{Q}_{r+1} = \boldsymbol{Q}_r \boldsymbol{P}_r$, ekkor $\boldsymbol{Q}_{r+1}$ unitér, mert két unitér mátrix szorzata unitér. Így a (4.36) egyenlőség $(r+1)$-re is igaz. Mivel a (4.36) egyenlőség $r = 1$ esetén igaz a 4.7-es tétel miatt, így minden r-re, $1 \leq r \leq n-1$, igaz. Ha $r = n-1$ a (4.36) egyenlőség jobb oldala felső háromszög mátrix, ezt $\boldsymbol{U}$-val jelölve és $\boldsymbol{Q}_{n-1}$-et $\boldsymbol{Q}$-val jelölve, kapjuk a tétel állítását. □

A tétel következményeit bizonyítás nélkül közöljük, a bizonyítások megtalálhatók [3] 80-82. oldalán.

Következmény. *Az n-ed rendű $\boldsymbol{A}$ négyzetes mátrixnak n darab sajátértéke van.*

A tétel következő következménye előtt bevezetjük a normálmátrix fogalmát.

Definíció. Az $\boldsymbol{A}$ mátrix normálmátrix, ha $\boldsymbol{A}^H \boldsymbol{A} = \boldsymbol{A} \boldsymbol{A}^H$.

Következmény. *Az n-ed rendű $\boldsymbol{A}$ négyzetes mátrix diagonális mátrixszá transzformálható egy unitér transzformációval akkor és csak akkor, ha az $\boldsymbol{A}$ mátrix normálmátrix.*

Következmény. *Ha az $\boldsymbol{A}$ Hermite-féle mátrix, akkor az $\boldsymbol{U}$ mátrix a (4.35) egyenlőségben valós diagonális mátrix, és ezért az $\boldsymbol{A}$ mátrix sajátértékei valósak.*

Következmény. *Ha az $\boldsymbol{A}$ mátrix valós, és minden sajátértéke valós, akkor az $\boldsymbol{A}$ mátrix valós felső háromszögmátrixszá alakítható egy ortogonális transzformációval.*

Következmény. *Ha az $\boldsymbol{A}$ mátrix valós és szimmetrikus, akkor az $\boldsymbol{A}$ mátrix valós diagonális mátrixszá alakítható egy ortogonális transzformációval.*

Megmutattuk, hogy az n-ed rendű $\boldsymbol{A}$ négyzetes mátrixnak n darab sajátértéke van, de ezek a sajátértékek nem feltétlen különbözőek. Most megvizsgáljuk hány darab különböző, azaz lineárisan független sajátvektora van egy mátrixnak. Kiemelt fontosságú, hogy egy n-ed rendű négyzetes mátrixnak van-e n darab lineárisan független sajátvektora. Ha igen, akkor ezek a lineárisan független sajátvektorok bázist alkotnak.

Definíció. Ha egy mátrix sajátvektorai nem alkotnak bázist, akkor a mátrixot defektív mátrixnak nevezzük.

Definíció. A mátrixot amelynek oszlopai egy nemdefektív mátrix sajátvektorai, modálmátrixnak nevezzük.

A modálmátrix nem egyértelmű. Egyrészt nincs meghatározva a mátrix oszlopainak sorrendje, másrészt a sajátvektorok tetszőlegesen méretezhetők.

4.9. Tétel. *Legyen $\boldsymbol{A}$ n-ed rendű mátrix és legyen m darab különböző sajátértéke, $m \leq n$. Az $\boldsymbol{A}$ mátrixnak legalább m darab lineárisan független sajátvektora van. [3, 86. oldal, Theorem 5.6]*

Bizonyítás. A tételt először az $\boldsymbol{U} = [u_{ij}]$ felső háromszögmátrixra bizonyítjuk. Jelölje λ_i az $\boldsymbol{U}$ mátrix sajátértékeit, λ_k pedig a többszörös sajátértékeit. Legyen

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{0} & \boldsymbol{U}_{22} \end{bmatrix}, \tag{4.40}$$

ahol $\boldsymbol{U}_{11}$ és $\boldsymbol{U}_{22}$ négyzetes mátrixok. Particionáljuk úgy az $\boldsymbol{U}$ mátrixot, hogy $\boldsymbol{U}_{11}$ mátrixnak ne legyen diagonális eleme λ_k, az $\boldsymbol{U}_{22}$ mátrixnak pedig legyen az első diagonális eleme λ_k. Tekintsük az $(\boldsymbol{U} - \lambda_k \boldsymbol{I}) = \boldsymbol{0}$ egyenlőséget, amit írhatunk az

$$\begin{bmatrix} \boldsymbol{U}_{11} - \lambda_k \boldsymbol{I} & \boldsymbol{U}_{12} \\ \boldsymbol{0} & \boldsymbol{U}_{22} - \lambda_k \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \tag{4.41}$$

vagy a

$$(\boldsymbol{U}_{11} - \lambda_k \boldsymbol{I})\boldsymbol{x}_1 = -\boldsymbol{U}_{12}\boldsymbol{x}_2 \tag{4.42}$$

és

$$(\boldsymbol{U}_{22} - \lambda_k \boldsymbol{I})\boldsymbol{x}_2 = \boldsymbol{0} \tag{4.43}$$

alakban. A particionálás miatt az $(\boldsymbol{U}_{22} - \lambda_k \boldsymbol{I})$ mátrix első oszlopa nulla, ezért $\boldsymbol{x}_2 = \boldsymbol{e}_1$, ahol $\boldsymbol{e}_1$ a megfelelő rendű egységmátrix első oszlopa. A particionálás miatt az

$(\boldsymbol{U}_{11} - \lambda_k \boldsymbol{I})\boldsymbol{x}_1$ nemszinguláris. Ekkor az (4.42) egyenlőségből

$$\boldsymbol{x}_1 = -(\boldsymbol{U}_{11} - \lambda_k \boldsymbol{I})^{-1}\boldsymbol{U}_{12}\boldsymbol{e}_2\,. \tag{4.44}$$

A (4.41) egyenlőségből látszik, hogy ha $\boldsymbol{U}_{11}$ $(r-1)$-ed rendű, akkor az $\boldsymbol{x}$ sajátvektor az r-edik oszlopa valamilyen n-ed rendű felső egység háromszögmátrixnak. Így az $\boldsymbol{U}$ mátrix m darab különböző sajátértékéhez tartozik különböző sajátvektor, amelyek egy felső egység háromszögmátrix különböző oszlopai. Mivel a felső egység háromszögmátrix nemszinguláris, a sajátvektorok lineárisan függetlenek.

Hogy a tételt az általános $\boldsymbol{A}$ mátrixra belássuk, legyen

$$\boldsymbol{Q}^H \boldsymbol{A}\boldsymbol{Q} = \boldsymbol{U}, \tag{4.45}$$

ahol $\boldsymbol{Q}$ unitér mátrix. Legyen $\boldsymbol{V}$ $n\times m$-es mátrix a $\lambda_1, \lambda_2, ..., \lambda_m$ különböző sajátértékekhez tartozó lineárisan független sajátvektorok mátrixa. Legyen $\boldsymbol{\Lambda}_1 = diag(\lambda_i)$ az m-ed rendű diagonális mátrix a $\lambda_1, \lambda_2, ..., \lambda_m$ különböző sajátértékkel. Ekkor

$$\boldsymbol{U}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}_1\,. \tag{4.46}$$

A (4.45) és a (4.46) egyenlőségekből következik, hogy

$$\boldsymbol{A}\boldsymbol{Q}\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{V}\boldsymbol{\Lambda}_1\,.$$

Azaz a $\boldsymbol{QV}$ mátrix oszlopai az $\boldsymbol{A}$ mátrixnak a $\lambda_1, \lambda_2, ..., \lambda_m$ különböző sajátértékeihez tartozó lineárisan független sajátvektorok mátrixa. Mivel $\boldsymbol{V}$ mátrix oszlopai lineárisan függetlenek és $\boldsymbol{Q}$ unitér mátrix, így a $\boldsymbol{QV}$ mátrix nemszinguláris, tehát a sajátvektorok lineárisan függetlenek. □

A tétel következményeit bizonyítás nélkül közöljük, a bizonyítások megtalálhatók [3] 88. oldalán.

Következmény. *Ha egy n-ed rendű mátrixnak n darab különböző sajátértéke van, akkor nem defektív mátrix.*

Következmény. *Ha egy mátrix sajátértéke egyszeres, a hozzátartozó sajátvektor egyértelmű, a méretezést leszámítva.*

Következmény. *Ha egy mátrix sajátértéke r-szeres, akkor legfeljebb r darab lineárisan független sajátvektor tartozik hozzá.*

A Iteratív eljárások fejezetben láttuk, hogy az iteratív eljárások konvergenciája összefügg a mátrixok normájával. Megvizsgálunk néhány, a sajátérték, a spektrális sugár és a mátrix norma közötti összefüggést.

Definíció. Legyen $\boldsymbol{A}$ n-ed rendű mátrix, λ_i sajátvektorokkal, $1 \leq i \leq n$. Az $\boldsymbol{A}$ mátrix spektrális sugara $\rho(\boldsymbol{A}) = \max_i |\lambda_i|$.

4.10. Tétel. *Legyen $\boldsymbol{A}$ négyzetes mátrix. Ekkor $\rho(\boldsymbol{A}) \leq \|\boldsymbol{A}\|$ tetszőleges normával. [3, 89. oldal, Theorem 5.7]*

Bizonyítás. Legyen $\boldsymbol{x}$ az $\boldsymbol{A}$ mátrix egy sajátvektora és legyen λ a hozzátartozó sajátérték, azaz $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. Így a normák tulajdonságai alapján

$$\|\lambda\boldsymbol{x}\| = \|\lambda\| \, \|\boldsymbol{x}\| = \|\boldsymbol{Ax}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{x}\| \, .$$

Mivel $\boldsymbol{x} \neq \boldsymbol{0}$, így $|\lambda| \leq \|\boldsymbol{A}\|$. □

4.11. Tétel. *Legyen $\boldsymbol{A}$ $m \times n$-es valós mátrix. Ekkor*

$$\|\boldsymbol{A}\|_2 = \left[\rho(\boldsymbol{A}^T\boldsymbol{A})\right]^{1/2} .$$

[3, 89. oldal, Theorem 5.8]

Bizonyítás. Az indukált mátrixnorma definíciójából,

$$\|\boldsymbol{A}\|_2^2 = \left(\max_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{Ax}\|_2}{\|\boldsymbol{x}\|_2}\right)^2 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \left(\frac{\|\boldsymbol{Ax}\|_2}{\|\boldsymbol{x}\|_2}\right)^2 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \left(\frac{\boldsymbol{x}\boldsymbol{A}^T\boldsymbol{Ax}}{\boldsymbol{x}^T\boldsymbol{x}}\right) . \tag{4.47}$$

Mivel az $\boldsymbol{A}^T\boldsymbol{A}$ mátrix valós és szimmetrikus, a sajátvektorai valósak és az $\boldsymbol{X}$ modálmátrixa ortogonális, a 4.8-as tétel következményei miatt. Legyen $\boldsymbol{\Lambda}$ a $\boldsymbol{A}^T\boldsymbol{A}$ mátrix sajátvektorainak diagonális mátrixa, azaz $\boldsymbol{A}^T\boldsymbol{AX} = \boldsymbol{X\Lambda}$. Legyen $\boldsymbol{x} = \boldsymbol{Xz}$. Mivel $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}$, következik, hogy $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{z}^T\boldsymbol{\Lambda z}$, és $\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{z}^T\boldsymbol{z}$. Így a (4.47) egyenlőség alakja

$$\|\boldsymbol{A}\|_2^2 = \max_{\boldsymbol{z}\neq\boldsymbol{0}} \left(\frac{\boldsymbol{z}^T\boldsymbol{\Lambda z}}{\boldsymbol{z}^T\boldsymbol{z}}\right) = \max_{\boldsymbol{z}\neq\boldsymbol{0}} \left[\left(\sum_{i=1}^n \lambda_i \boldsymbol{z}_i^2\right) / \left(\sum_{i=1}^n \boldsymbol{z}_i^2\right)\right] . \tag{4.48}$$

Legyen

$$\theta_i = \frac{\boldsymbol{z}_i^2}{\sum_{i=j}^n \boldsymbol{z}_j^2} \, ,$$

ezt a (4.48) egyenlőségbe helyettesítve kapjuk. hogy

$$\|\boldsymbol{A}\|_2^2 = \max \sum_i^n \theta_i \lambda_i \, ,$$

ahol $\theta \geq 0$ és $\sum_{i=1}^n \theta_i = 1$. A $\sum_i^n \theta_i\lambda_i$ kifejezést maximalizáljuk. A legnagyobb pozitív λ_i-hoz tartozó θ_i-t válasszuk egynek, az összes többi θ_i pedig nullának. Mivel az

$\boldsymbol{A}^T\boldsymbol{A}$ mátrixhoz tartozó összes sajátérték pozitív, a legnagyobb pozitív sajátértéke az $\boldsymbol{A}^T\boldsymbol{A}$ mátrixnak a spektrális sugara. Azaz $\|\boldsymbol{A}\|_2^2 = \rho(\boldsymbol{A}^T\boldsymbol{A})$. □

Következmény. *Ha* $\boldsymbol{A} = \boldsymbol{A}^T$, $\|\boldsymbol{A}\|_2 = \rho(\boldsymbol{A})$.

Bizonyítás. Ha $\boldsymbol{A} = \boldsymbol{A}^T$, akkor $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}^2$. Ha λ sajátértéke az $\boldsymbol{A}$ mátrixnak, akkor λ^2 sajátértéke az $\boldsymbol{A}^2$ mátrixnak. □

Következmény. $\|\boldsymbol{A}\|_2^2 \leq \|\boldsymbol{A}\|_1 \|\boldsymbol{A}\|_\infty$.

Bizonyítás. Legyen $\boldsymbol{x}$ az $\boldsymbol{A}^T\boldsymbol{A}$ mátrix $\rho\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ sajátértékéhez tartozó sajátvektora. Ekkor

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}\rho(\boldsymbol{A}^T\boldsymbol{A})\,,$$

és a tételt alkalmazva

$$\left\|\boldsymbol{x}\rho(\boldsymbol{A}^T\boldsymbol{A})\right\|_\infty = \|\boldsymbol{x}\|_\infty \|\boldsymbol{A}\|_2^2 = \left\|\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}\right\|_\infty \leq \left\|\boldsymbol{A}^T\right\|_\infty \|\boldsymbol{A}\|_\infty \|\boldsymbol{x}\|_\infty,$$

ahol $\|\boldsymbol{x}\|_\infty \neq 0$ és $\left\|\boldsymbol{A}^T\right\|_\infty = \|\boldsymbol{A}\|_1$. □

Definíció. Legyen $\boldsymbol{A} = [a_{ij}]$ n-ed rendű mátrix. Ekkor az $\boldsymbol{A}$ mátrix nyoma $tr(\boldsymbol{A}) = \sum_{i=1}^n a_{ii}$.

4.12. Tétel. *A mátrixhoz tartozó sajátértékek összege egyenlő a mátrix nyomával. [3, 91. oldal, Theorem 5.9]*

Bizonyítás. Bármilyen $\boldsymbol{B}$ és $\boldsymbol{C}$ mátrixokra, amelyeknek a $\boldsymbol{BC}$ és $\boldsymbol{CB}$ szorzatuk definiálva van, $tr(\boldsymbol{BC}) = tr(\boldsymbol{CB})$. A 4.8-as tételből tudjuk, hogy $\boldsymbol{Q}^H\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{U}$. Ezért

$$tr(\boldsymbol{U}) = tr((\boldsymbol{Q}^H\boldsymbol{A})\boldsymbol{Q}) = tr(\boldsymbol{Q}(\boldsymbol{Q}^H\boldsymbol{A})) = tr(\boldsymbol{A}),$$

az $\boldsymbol{U}$ mátrix nyoma pedig megegyezik az $\boldsymbol{U}$ mátrix sajátvektorainak összegével. □

4.4. Kovergens mátrixok

A különböző iteratív eljárások konvergenciája gyakran egy $\boldsymbol{A}^r$ mátrix viselkedésétől függ, miközben r tart a végtelenhez. Bizonyos iteratív eljárások akkor és csak akkor konvergensek, ha $\boldsymbol{A}^r$ a nullmátrixhoz konvergál, ahogy növekszik r. [3]

Definíció. Az n-ed rendű $\boldsymbol{A}$ mátrix konvergens, ha $\lim_{r\to\infty} \|\boldsymbol{A}^r\| = 0$.

4.13. Tétel. *Legyen $\boldsymbol{U}$ felső háromszögmátrix és $\rho(\boldsymbol{U}) < 1$. Ekkor létezik $\boldsymbol{B}$ diagonális mátrix, amelyre $\left\|\boldsymbol{B}\boldsymbol{U}\boldsymbol{B}^{-1}\right\|_\infty < 1$ és $\left\|\boldsymbol{B}\boldsymbol{U}\boldsymbol{B}^{-1}\right\|_1 < 1$. [3, 92. oldal, Theorem 5.10]*

Bizonyítás. Legyen $\boldsymbol{U} = [u_{ij}]$ n-ed rendű felső háromszögmátrix. Legyen $\boldsymbol{B} = diag(1, \beta, \beta^2, ..., \beta^{n-1})$, ahol válasszunk $\beta > 1$ skalárt úgy, hogy

$$\beta^{j-i} > \frac{(n-1)\,|u_{ij}|}{1-\rho(\boldsymbol{U})}\,, \qquad j > i \tag{4.49}$$

teljesüljön. Legyen $\boldsymbol{V} = [v_{ij}] = \boldsymbol{B}\boldsymbol{U}\boldsymbol{B}^{-1}$. Ekkor $j > i$-re $v_{ij} = u_{ij}/\beta^{j.i}$ és a (4.49) egyenlőtlenség miatt

$$|v_{ij}| < \frac{1-\rho(\boldsymbol{U})}{n-1}\,, \qquad j > i. \tag{4.50}$$

De $|v_{ii}| = |u_{ii}| < \rho(\boldsymbol{U})$ és $v_{ij} = 0$ minden $i > j$-re. Ezért

$$\|\boldsymbol{V}\|_\infty \leq \rho(\boldsymbol{U}) + \max_i \sum_{j=i+1}^{n} |v_{ij}|\,,$$

és így a (4.50) egyenlőtlenség miatt

$$\|\boldsymbol{V}\|_\infty \leq \rho(\boldsymbol{U}) + (n-1)\left[\frac{1-\rho(\boldsymbol{U})}{n-1}\right] = 1\,.$$

Hasonlóan, $\|\boldsymbol{V}\|_1 < 1$. □

Következmény. *A tétel feltételei mellett* $\left\|\boldsymbol{B}\boldsymbol{U}\boldsymbol{B}^{-1}\right\|_2 < 1$.

Bizonyítás. A 4.11-es tételből következik. □

A tétel fontossága, hogy ha az $\boldsymbol{A}$ mátrix spektrális sugara kisebb mint 1, akkor mindig létezik olyan norma, amiben az $\boldsymbol{A}$ mátrix normája kisebb mint 1.

4.14. Tétel. *Az n-ed rendű $\boldsymbol{A}$ mátrix konvergenciájának szükséges és elégséges feltétele, hogy $\rho(\boldsymbol{A}) < 1$. [3, 93. oldal, Theorem 5.11]*

Bizonyítás. A tételt először egy $\boldsymbol{U} = [u_{ij}]$ felső háromszögmátrixra bizonyítjuk, majd a 4.8-as tétellel kiterjesztjük a bizonyítást tetszőleges négyzetes mátrixra. Az i-edik diagonális eleme $\boldsymbol{U}^r$ mátrixnak u_{ii}^r, ezért $\boldsymbol{U}$ nem lehet konvergens, ha $u_{ii} \geq$ ≥ 1 bármilyen i-re. Ekkor ugyanis az $\boldsymbol{U}$ mátrix hatványozásakor az aktuális u_{ii}^r hatványok folyamatosan nőnek. Ha $u_{ii} \leq 1$, az $\boldsymbol{U}$ mátrix hatványozásakor az aktuális u_{ii}^r hatványok folyamatosan csökkennek. Mivel $\rho(\boldsymbol{U}) = \max_i |u_{ii}|$, ezért $\rho(\boldsymbol{U}) < 1$ szükséges feltétele a konvergenciának. Bizonyítjuk a feltétel elégségességét. A 4.13-es és a 4.11-es tétel következménye, hogy ha $\rho(\boldsymbol{U}) < 1$, akkor létezik olyan $\boldsymbol{B}$ diagonális mátrix, hogy $\boldsymbol{V} = \boldsymbol{B}\boldsymbol{U}\boldsymbol{B}^{-1}$ és $\|\boldsymbol{V}\|_2 < 1$. Ezért $\boldsymbol{U} = \boldsymbol{B}^{-1}\,\boldsymbol{V}\boldsymbol{B}$ és $\boldsymbol{U}^r = \boldsymbol{B}^{-1}\,\boldsymbol{V}^r\boldsymbol{B}$, mivel $\|\boldsymbol{B}\|_2 \left\|\boldsymbol{B}^{-1}\right\|_2 = \beta^{n-1}$. Ezekből következik, hogy $\|\boldsymbol{U}^r\|_2 \leq \beta^{n-1}\,\|\boldsymbol{V}\|_2^r$. Mivel $\|\boldsymbol{V}\|_2 < 1$ és $\lim_{r\to\infty} \|\boldsymbol{U}^r\| = 0$, az $\boldsymbol{U}$ mátrix konvergens.

Általános esetben, a 4.8-as tételből $\boldsymbol{Q}^H\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{U}$, ezért $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{U}\boldsymbol{Q}^H$ és $\boldsymbol{A}^r = = \boldsymbol{Q}\boldsymbol{U}^r\boldsymbol{Q}^H$. Mivel $\|\boldsymbol{Q}\|_2 \left\|\boldsymbol{Q}^H\right\|_2 = 1$, $\|\boldsymbol{A}^r\|_2 = \|\boldsymbol{U}^r\|_2$. Ebből az egyenlőségből következik, hogy a tétel általános esetben is igaz.

□

5. Felső relaxációs eljárás (SOR)

Stacionárius iterációs módszereknél minden olyan módosítás ami csökkenti a spektrális sugarat, meggyorsítja a konvergenciáját a lineáris egyenletrendszer megoldásának. A SOR módszer az egyik tagja annak a nagy módszercsaládnak, amelyet az iterációs eljárások konvergenciájának gyorsítására dolgoztak ki [14].

A SOR módszer konvergencia kritériumát szimmetrikus együtthatómátrixok esetén Alexander Ostrowski dolgozta ki 1954-ben [12]. Broyden elégséges konvergencia feltételeket adott 1964-ben, szimmetrikus és nemszimmetrikus együtthatómátrixok esetére is [2]. Ebben a fejezetben röviden bemutatjuk a SOR módszert, és összefoglaljuk Broyden a SOR módszer konvergenciájával kapcsolatos eredményeit [2].

5.1. Általános konvergencia eredmények

Legyen

$$\boldsymbol{M}\boldsymbol{x} = \boldsymbol{c}$$

lineáris egyenletrendszer, ahol $\boldsymbol{M}$ nemszinguláris, valós mátrix, $\boldsymbol{c}$ valós vektor, $\boldsymbol{x}$ ismeretlen vektor. Legyen

$$\boldsymbol{A} = \boldsymbol{D}\boldsymbol{M}, \qquad \boldsymbol{b} = \boldsymbol{D}\boldsymbol{c}, \tag{5.1}$$

ahol $\boldsymbol{D}$ diagonális mátrix. Így

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \tag{5.2}$$

A $\boldsymbol{D}$ mátrixot válasszuk úgy, hogy az $\boldsymbol{A}$ mátrix főátlójában lévő összes elem egységelem legyen. Ekkor az $\boldsymbol{A}$ mátrixot kifejezhetjük három mátrix összegeként,

$$\boldsymbol{A} = \boldsymbol{I} + \boldsymbol{L} + \boldsymbol{U}, \tag{5.3}$$

ahol $\boldsymbol{I}$ az egységmátrix, $\boldsymbol{L}$ alsó háromszögmátrix, $\boldsymbol{U}$ pedig felső háromszögmátrix.

Legyen $\boldsymbol{x}_{i+1}$ és $\boldsymbol{x}_i$ egymást követő közelítő megoldása az (5.2) egyenletrendszernek. Ekkor a SOR módszer szerint

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \omega\left[\left(\boldsymbol{I} + \boldsymbol{U}\right)\boldsymbol{x}_i + \boldsymbol{L}\boldsymbol{x}_{i+1} - \boldsymbol{b}\right], \tag{5.4}$$

ahol ω skalár. Az (5.3) egyenlőséget felhasználva, az (5.4) egyenlőségből az $\boldsymbol{U}$ mátrix eliminálásával kapjuk, hogy

$$(\boldsymbol{I} + \omega \boldsymbol{L})(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) = -\omega(\boldsymbol{A}\boldsymbol{x}_i - \boldsymbol{b}). \tag{5.5}$$

Definíció. Az $\boldsymbol{x}_i$ közelítő megoldáshoz tartozó maradéktag $\boldsymbol{\epsilon}_i = \boldsymbol{A}\boldsymbol{x}_i - \boldsymbol{b}$.

Ebből következik, hogy

$$\boldsymbol{\epsilon}_{i+1} = \boldsymbol{A}\boldsymbol{x}_{i+1} - \boldsymbol{b},$$

és

$$\boldsymbol{A}^{-1}(\boldsymbol{\epsilon}_{i+1} - \boldsymbol{\epsilon}_i) = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i.$$

Így az (5.5) egyenlőségből

$$(\boldsymbol{I} + \omega \boldsymbol{L})\boldsymbol{A}^{-1}(\boldsymbol{\epsilon}_{i+1} - \boldsymbol{\epsilon}_i) = -\omega \boldsymbol{\epsilon}_i. \tag{5.6}$$

Az (5.6) egyenlőségből

$$\boldsymbol{\epsilon}_{i+1} = \left[\boldsymbol{I} - \omega \boldsymbol{A}\left(\boldsymbol{I} + \omega \boldsymbol{L}\right)^{-1}\right] \epsilon_i \,. \tag{5.7}$$

Írjuk ezt az egyenlőséget az

$$\boldsymbol{\epsilon}_{i+1} = (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\epsilon}_i$$

egyszerűbb formában. [2]

5.1. Tétel. *Azoknál az iteratív eljárásoknál, ahol a közelítő megoldások maradéktagja kifejezhető az*

$$\boldsymbol{\epsilon}_{i+1} = (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\epsilon}_i$$

alakban, a konvergencia elégséges feltétele, hogy létezzen olyan $\boldsymbol{S}$ és $\boldsymbol{G}$ mátrix, hogy

$$\boldsymbol{S} > 0$$

és

$$\boldsymbol{G} = \boldsymbol{B}^T\boldsymbol{S} + \boldsymbol{S}\boldsymbol{B} - \boldsymbol{B}^T\boldsymbol{S}\boldsymbol{B} > 0.$$

[2, 137. oldal, Theorem 1]

Bizonyítás. Legyen $f_i = \boldsymbol{\epsilon}_i^T \boldsymbol{S} \boldsymbol{\epsilon}_i$. Mivel a hipotézis szerint $\boldsymbol{S} > 0$, $f_i > 0$, minden $\epsilon_i \neq 0$ esetén. A konvergencia szükséges és elégséges feltétele, ha $f_i \to 0$, miközben $i \to \infty$.

$$\begin{aligned} f_{i+1} &= \boldsymbol{\epsilon}_{i+1}^T \boldsymbol{S} \boldsymbol{\epsilon}_{i+1} \\ &= \boldsymbol{\epsilon}_i^T (\boldsymbol{I} - \boldsymbol{B}^T) \boldsymbol{S} (\boldsymbol{I} - \boldsymbol{B}) \boldsymbol{\epsilon}_i \\ &= \boldsymbol{\epsilon}_i^T (\boldsymbol{S} - \boldsymbol{G}) \boldsymbol{\epsilon}_i \,. \end{aligned}$$

Legyen $\phi_i = \boldsymbol{\epsilon}_i^T \boldsymbol{G} \boldsymbol{\epsilon}_i$, ekkor $f_{i+1} = f_i - \phi_i$. Egy elégséges feltétele annak, hogy $f_i \to 0$, miközben $i \to \infty$, ha létezik olyan k konstans, hogy

$$\phi_i \geq k f_i \,, \tag{5.8}$$

mert ekkor $f_{i+1} \leq (1-k) f_i$ és az f_i sorozat konvergál.

Az $\boldsymbol{S}$ mátrix minden sajátértéke valós és pozitív, mert $\boldsymbol{S} > 0$ (4.5-ös tétel). Ha $\lambda_{\max}$ az $\boldsymbol{S}$ mátrix legnagyobb sajátértéke, akkor

$$f_i \leq \lambda_{\max} \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i \,. \tag{5.9}$$

Mivel $\boldsymbol{S} = \boldsymbol{S}^T$, a $\boldsymbol{G}$ mátrix szimmetrikus és a sajátértékei valósak. Legyen a $\boldsymbol{G}$ mátrix legkisebb sajátértéke $\mu_{\min}$. Ekkor $\phi_i \geq \mu_{\min} \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i$. Az (5.9) egyenlőségből

$$\phi_i \geq \frac{\mu_{\min}}{\lambda_{\max}} f_i \,.$$

Ha $\boldsymbol{G} > 0$, akkor $\mu_{\min} > 0$, és teljesül az (5.8) feltétel. □

Következmény. *Ha $\boldsymbol{S} > 0$ és $\boldsymbol{G} \leq 0$, azaz ha $\boldsymbol{S}$ mátrix pozitív definit és a $\boldsymbol{G}$ mátrix negatív szemidefinit, akkor az iteratív eljárás sohasem konvergál.*

Most az 5.1-es tételt alkalmazzuk a SOR módszerre. Az 5.7 egyenlőség miatt, $\boldsymbol{B} = \omega \boldsymbol{A} (\boldsymbol{I} + \omega \boldsymbol{L})^{-1}$, ezért a SOR módszer konvergenciájának egy elégséges feltétele, hogy létezik olyan $\boldsymbol{S} > 0$ mátrix, hogy

$$\omega (\boldsymbol{I} + \omega \boldsymbol{L}^T)^{-1} \boldsymbol{A}^T \boldsymbol{S} + \omega \boldsymbol{S} \boldsymbol{A} (\boldsymbol{I} + \omega \boldsymbol{L})^{-1} - \omega^2 (\boldsymbol{I} + \omega \boldsymbol{L}^T)^{-1} \boldsymbol{A}^T \boldsymbol{S} \boldsymbol{A} (\boldsymbol{I} + \omega \boldsymbol{L})^{-1} > 0. \tag{5.10}$$

Ezt a feltételt egyszerűsíthetjük a következő tétel felhasználásával.

5.2. Tétel. *A $\boldsymbol{P} > 0$ szükséges és elégséges feltétele, hogy $\boldsymbol{Q}^T \boldsymbol{P} \boldsymbol{Q} > 0$, ahol $\boldsymbol{Q}$ tetszőleges nemszinguláris mátrix. [2, 138. oldal, Lemma]*

Bizonyítás. Legyen $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{z}$. Ekkor $\boldsymbol{x}^T \boldsymbol{Q}^T \boldsymbol{P} \boldsymbol{Q} \boldsymbol{x} = \boldsymbol{z}^T \boldsymbol{P} \boldsymbol{z}$. Mivel $\boldsymbol{Q}$ mátrix nemszinguláris, minden nem nulla $\boldsymbol{x}$ vektorra létezik nem nulla $\boldsymbol{z}$ vektor, és fordítva. □

Mivel $(\boldsymbol{I} + \omega \boldsymbol{L})$ nemszinguláris, ezért az (5.10) elégséges feltétel a rövidebb

$$\omega \left[\boldsymbol{A}^T \boldsymbol{S} (\boldsymbol{I} + \omega \boldsymbol{L}) + (\boldsymbol{I} + \omega \boldsymbol{L}^T) \boldsymbol{S} \boldsymbol{A} - \omega \boldsymbol{A}^T \boldsymbol{S} \right] > 0 \tag{5.11}$$

formában írható [2].

5.2. Szimmetrikus mátrixok konvergenciája

Tegyük fel, hogy az $\boldsymbol{M}$ mátrix szimmetrikus. Legyen $\boldsymbol{S} = \boldsymbol{D}^{-1}\boldsymbol{M}^{-1}\boldsymbol{D}^{-1}$. Ekkor az (5.10) feltétel alakja,

$$\omega \left[\boldsymbol{D}^{-1}(\boldsymbol{I} + \omega\boldsymbol{L}) + (\boldsymbol{I} + \omega\boldsymbol{L}^T)\boldsymbol{D}^{-1} - \omega\boldsymbol{M}\right] > 0. \tag{5.12}$$

Az (5.1) és (5.3) egyenlőségekből, és $\boldsymbol{M}$ feltételezett szimmetrikussága miatt,

$$\boldsymbol{M} = \boldsymbol{D}^{-1}(\boldsymbol{I} + \boldsymbol{L} + \boldsymbol{U}) = (\boldsymbol{I} + \boldsymbol{L}^T + \boldsymbol{U}^T)\boldsymbol{D}^{-1}. \tag{5.13}$$

Az $\boldsymbol{M}$ mátrix az (5.13) egyenlőség szerinti két reprezentációjának felső háromszög partícióit egyenlővé téve kapjuk, hogy

$$\boldsymbol{D}^{-1}\boldsymbol{U} = \boldsymbol{L}^T\boldsymbol{D}^{-1}. \tag{5.14}$$

Így az (5.11) feltétel az

$$\omega(2 - \omega)\boldsymbol{D}^{-1} > 0 \tag{5.15}$$

alakra egyszerűsödik. Ha $\boldsymbol{M} > 0$ akkor az 5.2-es tételből következik, hogy $\boldsymbol{S} > 0$, és mivel $\boldsymbol{M} > 0$, ezért $\boldsymbol{D} > 0$, és az (5.15) feltétel teljesül. Ezért a SOR módszer konvergálni fog, ha $\boldsymbol{M} > 0$ és $0 < \omega < 2$. Azonban ha $\omega < 0$, vagy $\omega > 2$, a $\boldsymbol{G}$ mátrix negatív definitté válik, és a 5.1-es tétel következménye miatt a SOR módszer nem fog konvergálni. [2]

5.3. Nemszimmetrikus mátrixok konvergenciája

Legyen $\boldsymbol{S} = (\boldsymbol{A}^T)^{-1}\boldsymbol{A}^{-1}$ az (5.11) feltételben. Az $\boldsymbol{A}$ mátrix nemszinguláris, ezért az $(\boldsymbol{A}^T)^{-1}\boldsymbol{A}^{-1}$ mátrix pozitív definit. Ekkor a konvergencia feltétele,

$$\omega\boldsymbol{A}^{-1}(\boldsymbol{I} + \omega\boldsymbol{L}) + \omega(\boldsymbol{I} + \omega\boldsymbol{L}^T) + (\boldsymbol{A}^T)^{-1} - \omega^2\boldsymbol{I} > 0.$$

Az $\boldsymbol{A}$ mátrix nemszinguláris, ezért az 5.2-es tételt felhasználva, a feltétel a

$$\omega(\boldsymbol{I} + \omega\boldsymbol{L})\boldsymbol{A}^T + \omega\boldsymbol{A}(\boldsymbol{I} + \omega\boldsymbol{L}^T) - \omega^2\boldsymbol{A}\boldsymbol{A}^T > 0$$

alakra egyszerűsödik. Az $\boldsymbol{A}$ mátrixot felbontva (5.3) szerint és egyszerűsítve,

$$\omega(\boldsymbol{A} + \boldsymbol{A}^T) - \omega^2\left[(\boldsymbol{I} + \boldsymbol{U})(\boldsymbol{I} + \boldsymbol{U}^T) - \boldsymbol{L}\boldsymbol{L}^T\right] > 0. \tag{5.16}$$

Legyen $\boldsymbol{S} = \boldsymbol{I}$ az (5.11) feltételben. Ekkor a konvergencia elégséges feltétele

$$\omega \boldsymbol{A}^T(\boldsymbol{I} + \omega \boldsymbol{L}) + \omega(\boldsymbol{I} + \omega \boldsymbol{L}^T)\boldsymbol{A} - \omega^2 \boldsymbol{A}^T \boldsymbol{A} > 0,$$

az $\boldsymbol{A}$ mátrixot felbontva (5.3) szerint és egyszerűsítve,

$$\omega(\boldsymbol{A} + \boldsymbol{A}^T) - \omega^2 \left[(\boldsymbol{I} + \boldsymbol{U}^T)(\boldsymbol{I} + \boldsymbol{U}) - \boldsymbol{L}^T \boldsymbol{L}\right] > 0. \tag{5.17}$$

Megmutatjuk, hogy ha $\boldsymbol{A} + \boldsymbol{A}^T > 0$, akkor létezik olyan pozitív ω, ami kielégíti az (5.16) és az (5.17) feltételeket. [2]

5.3. Tétel. *Ha $\boldsymbol{P} > 0$ és $\boldsymbol{Q} = \boldsymbol{Q}^T$, akkor létezik olyan pozitív ω, hogy $\boldsymbol{P} + \omega \boldsymbol{Q} > 0$. [2, 139. oldal, Theorem 2]*

Bizonyítás. Legyen $f_1 = \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}$. Mivel $\boldsymbol{P} > 0$, minden sajátértéke valós és pozitív (4.5-ös tétel). Jelölje $\lambda_{\min}$ a $\boldsymbol{P}$ mátrix legkisebb sajátértékét. Ekkor

$$f_1 \geq \lambda_{\min} \boldsymbol{x}^T \boldsymbol{x}.$$

A $\boldsymbol{Q}$ mátrix szimmetrikus, ezért minden sajátértéke valós. Jelölje a $\boldsymbol{Q}$ mátrix legkisebb sajátértékét $\mu_{\min}$. Ha

$$f_2 = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x},$$

akkor

$$f_2 \geq \mu_{\min} \boldsymbol{x}^T \boldsymbol{x}.$$

Ha

$$f = \boldsymbol{x}^T (\boldsymbol{P} + \omega \boldsymbol{Q}) \boldsymbol{x},$$

akkor

$$f \geq (\lambda_{\min} + \omega \mu_{\min}) \boldsymbol{x}^T \boldsymbol{x}.$$

Az első eset, ha

$$\mu_{\min} \geq 0.$$

Ekkor $f > 0$ minden $\omega \geq 0$ esetén. A második eset, ha

$$\mu_{\min} = -\left|\mu_{\min}\right| < 0.$$

Ekkor

$$f > (\lambda_{\min} - \omega \left|\mu_{\min}\right|) \boldsymbol{x}^T \boldsymbol{x},$$

ezért $f > 0$ minden $\omega < \frac{\lambda_{\min}}{|\mu_{\min}|}$ és $\boldsymbol{x} \neq \boldsymbol{0}$ esetén. □

Elégséges konvergencia feltételt vezethetünk le az (5.17) feltételből is [2], legyen

$$\boldsymbol{P} = (\boldsymbol{I} + \boldsymbol{L}^T)(\boldsymbol{I} + \boldsymbol{L}) - \boldsymbol{U}^T\boldsymbol{U},$$

$$\boldsymbol{Q} = (\boldsymbol{I} + \boldsymbol{U}^T)(\boldsymbol{I} + \boldsymbol{U}) - \boldsymbol{L}^T\boldsymbol{L}.$$

Ekkor $\boldsymbol{P} + \boldsymbol{Q} = \boldsymbol{A} + \boldsymbol{A}^T$. Az (5.17) feltétel alakja

$$\omega^2\boldsymbol{P} - \omega(\omega - 1)\boldsymbol{Q} > 0. \tag{5.18}$$

Ezért a SOR módszer konvergálni fog, ha $\omega = 1$ és $\boldsymbol{P} > 0$. [2]

5.4. Konklúzió

Az 5.3-as tétel szerint az (5.16) és (5.17) feltételek teljesülnek, ha $\boldsymbol{A} + \boldsymbol{A}^T$ pozitív definit, és megfelelően kicsi ω-át választunk. Minden $\boldsymbol{A}$ mátrix kifejezhető szimmetrikus és antiszimmetrikus mátrixok összegeként. Ezért ha az $\boldsymbol{A}$ mátrixot így bontjuk fel, és a szimmetrikus mátrix pozitív definit, mindig lehetséges lesz olyan ω-t találni, hogy a SOR módszer konvergáljon. [2]

Az (5.18) feltétel más jellegű. Azt mutatja meg, hogy a SOR módszer konvergálni fog $\omega = 1$ esetén, ha

$$(\boldsymbol{I} + \boldsymbol{L}^T)(\boldsymbol{I} + \boldsymbol{L}) - \boldsymbol{U}^T\boldsymbol{U} > 0. \tag{5.19}$$

Ebből következik, hogy léteznek olyan mátrixok, amelyeknél az alsó háromszög domináltság fontos szempont a SOR módszer konvergenciájánál. A szélsőséges esetben, amikor $\boldsymbol{U}$ nulla lesz, az (5.19) feltétel fennáll, és ha ω értéke 1, az egyenletek egy lépésben megoldódnak. [2]

Azt, hogy a $\boldsymbol{P} > 0$ és az $\boldsymbol{A} + \boldsymbol{A}^T > 0$ feltételek nem ekvivalensek, legjobban két példával lehet szemléltetni. Legyen

$$\boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix}.$$

Ebben az esetben az $\boldsymbol{A}$ mátrix erősen antiszimmetrikus, és a szimmetrikus komponense pozitív definit, habár sem $\boldsymbol{P}$, sem $\boldsymbol{Q}$ nem az. A SOR módszer megfelelően kicsi ω esetén konvergál. Pl. $\omega = \frac{1}{4}$. [2]

Legyen

$$\boldsymbol{A} = \begin{bmatrix} 1 & \frac{1}{3} \\ 2 & 1 \end{bmatrix}.$$

Az $\boldsymbol{A}$ mátrix alsó háromszög dominált. A $\boldsymbol{P}$ mátrix pozitív definit, de $\boldsymbol{A} + \boldsymbol{A}^T$ és $\boldsymbol{Q}$ nem az. A konvergencia $\omega = 1$ esetén garantált. [2]

Érdemes kiemelni, hogy az (5.16), (5.17), (5.18) feltételek elégségesek, de nem szükségesek. A (5.16), (5.17) feltételeket nem kielégítő, nagyobb ω, nem feltétlenül jelenti, hogy a SOR módszer divergálni fog. Viszont a feltételekből látszik, hogy létezik kettő igen különböző típusa a nemszimmetrikus mátrixoknak, melyekre a SOR módszer mindig konvergál, ha megfelelő ω értéket választunk. [2]

6. A konjugált gradiens módszerek új rendszertana

6.1. A témakör felvezetése

Az egyik legjobban használható módszer az

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{6.1}$$

egyenletrendszer megoldására, ahol $\boldsymbol{A}$ valós nemszinguláris ritka mátrix, $\boldsymbol{b}$ pedig valós vektor, a konjugált gradiens módszer és az ebből származtatott különböző módszerek [5]. Hestenes és Siefel eredeti 1952-es módszere [9] csak akkor alkalmazható, ha az $\boldsymbol{A}$ együtthatómátrix szimmetrikus és pozitív definit. Az eredeti módszer óta már rengetek származtatott módszer született, amelyek nem csak szimmetrikus indefinit együtthatómátrixok esetén, de nemszimmetrikus együtthatómátrixok esetén is alkalmazhatóak [5]. A módszerek áttekintése azonban nehézkes lehet, például amiatt, hogy az algoritmusokat a különböző szerzők különböző módokon származtatják. Broyden 1996-os cikkében [5] rendszerezi a konjugált gradiens módszereket. Cikke támaszkodik más szerzők azonos célú korábbi munkáira, például S.F. Ashby, T.A. Manteuffel és P.E. Saylor 1990-es cikkére [1]. Ők minden algoritmust három mátrixszal jellemeznek. Az $\boldsymbol{A}$ együtthatómátrixszal, egy Hermit-féle pozitív definit $\boldsymbol{B}$ mátrixszal (a belső szorzat mátrixszal), és egy további $\boldsymbol{C}$ mátrixszal (a prekondicionáló mátrixszal). Cikkük erősen támaszkodik Faber és Manteuffel cikkére [8], akik a valós $\boldsymbol{M}$ mátrixot B-normális(s)-nek nevezik, ha a $\boldsymbol{B}$ mátrix Hermit-féle, pozitív definit, és

$$\boldsymbol{M}^T\boldsymbol{B} = \boldsymbol{B}p(\boldsymbol{M})\,, \tag{6.2}$$

ahol s a legkisebb fokszáma a $p(\boldsymbol{M})$ mátrixpolinomnak, amelyre a (6.2) egyenlőség teljesül [8]. Megmutatták, hogy egy 3-tagú rekurrens prekondicionált konjugált gradiens módszernél, ha a $\boldsymbol{C}\boldsymbol{A}$ mátrixszorzat B-normális(1), akkor a módszer véges lépésben leáll [8]. Broyden rendszertanában egy $\boldsymbol{G}$ Hesse-mátrixból és egy további $\boldsymbol{K}$ mátrixból származtatja a módszereket [5]. Broyden $\boldsymbol{G}$ mátrixa általában azonos az Ashby-féle $\boldsymbol{B}$ mátrixszal, azzal a különbséggel, hogy a pozitív definitség a $\boldsymbol{G}$

mátrixnál nem megkövetelt [5]. További összefüggés a két rendszertan között, hogy

$$\boldsymbol{CA} = \boldsymbol{KG} \tag{6.3}$$

ahol $\boldsymbol{C}$ és $\boldsymbol{A}$ az Ashby féle $\boldsymbol{C}$ és $\boldsymbol{A}$ mátrix [5]. Ebből következik, hogy a $\boldsymbol{KG}$ mátrixszorzatnak B-normális(1)-nek kell lennie. Legyen $\boldsymbol{KG}$ az $\boldsymbol{M}$ mátrix, és $\boldsymbol{G}$ a $\boldsymbol{B}$ mátrix a (6.2) egyenlőségben. Ekkor a B-normalitás követelménye

$$\boldsymbol{GK}^T\boldsymbol{G} = \boldsymbol{G}p(\boldsymbol{KG}), \tag{6.4}$$

ahol $p(\boldsymbol{KG})$ lineáris mátrixpolinom [5]. Ennek elégséges feltétele, ha a $\boldsymbol{K}$ mátrix szimmetrikus, ekkor $p(\boldsymbol{KG})$ helyett vehetjük a $\boldsymbol{KG}$ mátrixszorzatot [5]. Habár a $\boldsymbol{K}$ mátrix szimmetrikussága erősebb feltétel, mint a B-normalitás, de ahhoz még elég általános, hogy a standard módszereket is tartalmazza a rendszertan [5]. Ennek a mátrixfelbontásnak még egy következménye, hogy a hagyományos 2-tagú konjugált gradiens módszereknél, a két tipikus numerikus instabilitása ezeknek a módszereknek (lásd később), a $\boldsymbol{G}$ és a $\boldsymbol{K}$ mátrixokból könnyen megállapítható [5]. A $\boldsymbol{K}$ és $\boldsymbol{G}$ mátrixokon kívűl még szükségünk van bizonyos vektorokra, a generátor vektorokra, amelyek minden iterációval változnak. Ezeket kétféleképpen választhatjuk, az egyik választás vezet a 2-tagú (Hestenes-Stiefel) rekurrens módszerekhez, míg a másik választás a 3-tagú rekurrens (Lánczos) módszerekhez [5].

Legyen a $\phi(\boldsymbol{x})$ kvadratikus függvény

$$\phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{G}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{h}, \tag{6.5}$$

ahol $\boldsymbol{G}$ n-ed rendű szimmetrikus mátrix, de nem feltétlen pozitív definit, $\boldsymbol{h}$ pedig $\boldsymbol{n}$-ed rendű vektor. Legyen $\boldsymbol{x}_1$ a kezdeti értéke az $\boldsymbol{x}$ vektornak, és legyen $\boldsymbol{S}_i$ $n\times i$-s, vagy $n\times 2i$-s mátrix. Ekkor, ha $\boldsymbol{x}_{i+1}$ jelöli azt az $\boldsymbol{x}$ vektor értéket, amelyre a $\phi(\boldsymbol{x})$ függvény stacionárius az

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{S}_i\boldsymbol{z} \tag{6.6}$$

hipersík felett, ahol z i-ed vagy $2i$-ed rendű vektor, akkor

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_1 - \boldsymbol{S}_i(\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{S}_i)^{-1}\boldsymbol{S}_i^T\boldsymbol{g}_1\,, \tag{6.7}$$

ahol $\boldsymbol{g}_1$ a gradiensvektora a ϕ függvénynek az $\boldsymbol{x}_1$ helyen. Továbbá, ha

$$\boldsymbol{S}_i = \begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2 & \ldots & \boldsymbol{P}_i\end{bmatrix}\,, \tag{6.8}$$

ahol $\boldsymbol{P}_j$, $1 \le j \le i$, részmátrixok, egy vagy két oszloppal, úgy, hogy a $\boldsymbol{C}_j = \boldsymbol{P}_j^T\boldsymbol{G}\boldsymbol{P}_j$

mátrix nemszinguláris, és

$$\boldsymbol{P}_j^T \boldsymbol{G} \boldsymbol{P}_k = \boldsymbol{0}, \tag{6.9}$$

ahol $j \neq k$, $1 \leq j, k \leq i$. Az $\boldsymbol{x}_{i+1}$ vektor kifejezhető úgy is, hogy

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \boldsymbol{P}_i (\boldsymbol{P}_i^T \boldsymbol{G} \boldsymbol{P}_i)^{-1} \boldsymbol{P}_i^T \boldsymbol{g}_i \,, \tag{6.10}$$

ahol

$$\boldsymbol{g}_i = \boldsymbol{G}\boldsymbol{x}_i - \boldsymbol{h}. \tag{6.11}$$

A (6.7) egyenlőséget a $\boldsymbol{G}$ mátrixszal szorozva, és a $\boldsymbol{h}$ vektort mindkét oldalból kivonva,

$$\boldsymbol{g}_{i+1} = \boldsymbol{Q}_i \boldsymbol{g}_1 \,, \tag{6.12}$$

ahol

$$\boldsymbol{Q}_i = \boldsymbol{I} - \boldsymbol{G}\boldsymbol{S}_i (\boldsymbol{S}_i^T \boldsymbol{G} \boldsymbol{S}_i)^{-1} \boldsymbol{S}_i^T \tag{6.13}$$

[5].

Ahhoz, hogy ki tudjuk számolni az $\boldsymbol{x}_{i+2}$ vektort, szükség van a $\boldsymbol{P}_{i+1}$ mátrixra. Könnyen látható, hogy ha

$$\boldsymbol{P}_{i+1} = \boldsymbol{Q}_i^T \boldsymbol{W}_{i+1} \,, \tag{6.14}$$

akkor a (6.9) egyenlőség fennáll $(i+1)$-re, tetszőleges $\boldsymbol{W}_{i+1}$ mátrixszal. A $\boldsymbol{Q}_i$ mátrix túl nagy és sűrű, de mivel a $\boldsymbol{W}_{i+1}$ mátrix tetszőlegesen választható, ezzel egyszerűsíthetjük a (6.14) összefüggést. Két lehetséges választása a $\boldsymbol{W}_{i+1}$ mátrixnak különösen előnyös. [5]

Az első lehetőség, amikor a $\boldsymbol{G}$ mátrixnak nincs semmilyen különösebb blokk struktúrája, és amikor a $\boldsymbol{P}_j$ mátrixok $\boldsymbol{p}_j$ vektorok, a kapcsolódó $\boldsymbol{W}_j$ generátor mátrixok pedig $\boldsymbol{w}_j$ vektorok. Megmutatjuk, hogy ha

$$\boldsymbol{w}_j = \boldsymbol{K}\boldsymbol{g}_j \,, \tag{6.15}$$

ahol $\boldsymbol{K}$ tetszőleges szimmetrikus mátrix, akkor a (6.14) egyenlőség a

$$\boldsymbol{p}_{i+1} = \boldsymbol{K}\boldsymbol{g}_{i+1} - \boldsymbol{p}_i \alpha_i \tag{6.16}$$

alakra egyszerűsödik, ahol α_i konstans. [5]

Először bebizonyítjuk a következő tételt.

6.1. Tétel. *Legyen $\boldsymbol{x}_{i+1}$ stacionárius pontja a $\phi(x)$ függvénynek a (6.6) egyenlőségben definiált hipersík felett. Legyen $\boldsymbol{g}_{i+1}$ az $\boldsymbol{x}_{i+1}$ ponthoz tartozó gradiense a $\phi(\boldsymbol{x})$ függvénynek, és legyenek a $\boldsymbol{w}_j$ vektorok az $\boldsymbol{S}_i$ mátrix generátorai. Ekkor $\boldsymbol{w}_j^T \boldsymbol{g}_{i+1} = 0$, $i \leq j \leq i$. [5, 9. oldal, Lemma 1]*

Bizonyítás. A (6.12) egyenlőségből $\boldsymbol{w}_j^T \boldsymbol{g}_{i+1} = \boldsymbol{w}_j^T \boldsymbol{Q}_i \boldsymbol{g}_1$, a (6.13) és a (6.9) egyenlő-

ségből $\boldsymbol{Q}_{j-1}\boldsymbol{Q}_i = \boldsymbol{Q}_i$, $0 \leq j-1 \leq i$, ezért $\boldsymbol{w}_j^T\boldsymbol{g}_{i+1} = \boldsymbol{w}_j^T\boldsymbol{Q}_{j-1}\boldsymbol{Q}_i\boldsymbol{g}_1$. Ezért a (6.14) egyenlőségből $\boldsymbol{w}_j^T\boldsymbol{g}_{i+1} = \boldsymbol{p}_j\boldsymbol{Q}_i\boldsymbol{g}_1$. A tétel következik a (6.8) egyenlőségből, mivel a (6.13) egyenlőség miatt $\boldsymbol{S}_i^T\boldsymbol{Q}_i = 0$. □

Következmény. *Ha* $\boldsymbol{w_j} = \boldsymbol{K}\boldsymbol{g}_j$, $1 \leq j \leq i$, *akkor* $\boldsymbol{g}_j^T\boldsymbol{K}\boldsymbol{g}_k = 0$, $j \neq k$, $j \geq 1$, $k \leq i+1$.

Megmutatjuk, hogy ha a $\boldsymbol{K}$ mátrix definit, akkor a generátor vektorok fenti választásával

$$\boldsymbol{p}_j^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{g}_{i+1} = 0, \tag{6.17}$$

ahol $1 \leq j \leq i-1$. A (6.10) egyenlőségben i-t j-re cserélve, az egyenlőséget megszorozva a $\boldsymbol{KG}$ mátrixszal, és a $\boldsymbol{Kh}$ vektort kivonva mindkét oldalból, a (6.11) egyenlőségből kapjuk, hogy

$$\boldsymbol{K}\boldsymbol{g}_{j+1} = \boldsymbol{K}\boldsymbol{g}_j - \boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_j\gamma_j\,, \tag{6.18}$$

ahol $\gamma_j = \boldsymbol{p}_j^T\boldsymbol{g}_j/\boldsymbol{p}_j^T\boldsymbol{G}\boldsymbol{p}_j$. A $\boldsymbol{P}_j$ mátrix most egy vektor, ezért $\boldsymbol{p}_j$-vel jelöltük. Megszorozva ezt az egyenlőséget a $\boldsymbol{g}_{i+1}^T$ vektorral, a 6.1-es tétel következményéből kapjuk, hogy $\gamma_j\boldsymbol{p}_j^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{g}_{i+1} = 0$, $1 \leq j \leq i-1$. A (6.18) egyenlőséget megszorozva a $\boldsymbol{g}_{j+1}^T$ vektorral, a 6.1-es tétel következménye, hogy ha $\boldsymbol{g}_{j+1} \neq \boldsymbol{0}$, és minthogy $\boldsymbol{K}$ definit, akkor $\gamma \neq 0$, így kapjuk a (6.17) egyenlőséget [5].

A (6.13) és a (6.14) egyenlőségből

$$\boldsymbol{p}_{i+1} = \boldsymbol{K}\boldsymbol{g}_{i+1} - \boldsymbol{S}_i(\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{S}_i)^{-1}\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{g}_{i+1}\,, \tag{6.19}$$

de a (6.17) egyenlőség miatt ennek az egyenlőségnek csak a $\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{g}_{i+1}$ tagja nem nulla, ebből következik a (6.16) egyenlőség és az $\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{S}_i$ mátrix diagonalitása [5].

Az α_i konstansot úgy választjuk a (6.16) egyenlőségben, hogy $\boldsymbol{p}_i^T\boldsymbol{G}\boldsymbol{p}_{i+1} = 0$ legyen. Ez mindig lehetséges, ha $\boldsymbol{p}_i^T\boldsymbol{G}\boldsymbol{p}_i \neq 0$. Ez a feltétel mindig teljesül a nem nulla $\boldsymbol{p}_i$ vektorokra, ha a $\boldsymbol{G}$ mátrix pozitív definit. Az olyan algoritmusokat, ahol a $\boldsymbol{G}$ mátrix pozitív definit, b-stabilisnak nevezzük. Ha a $\boldsymbol{K}$ mátrix definit, egy másféle numerikus instabilitást kerülhetünk el, amikor is az egymást követő lépések túl kicsik lesznek [5]. Azokat az algoritmusokat, ahol a $\boldsymbol{K}$ mátrix definit, ω-stabilisnak nevezzük. Így a konjugált gradiens algoritmusok kétféle numerikus instabilitása a $\boldsymbol{G}$ és a $\boldsymbol{K}$ mátrixoktól függenek [5].

A második lehetőség a $\boldsymbol{w}_j$ vektor választására, ha

$$\boldsymbol{w}_j = \boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_{j-1}\,, \tag{6.20}$$

ahol a $\boldsymbol{p}_1$ vektort tetszőlegesen választhatjuk és $\boldsymbol{p}_0 = \boldsymbol{0}$. Ezzel a választással a (6.14)

egyenlőség alakja

$$\boldsymbol{p}_{i+1} = \boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_i - \boldsymbol{p}_i\alpha_i - \boldsymbol{p}_{i-1}\beta_{i-1}\,, \tag{6.21}$$

ahol az α_i és a β_{i-1} konstansokat úgy választjuk, hogy

$$\boldsymbol{p}_{i-1}^T\boldsymbol{G}\boldsymbol{p}_{i+1} = \boldsymbol{p}_i^T\boldsymbol{G}\boldsymbol{p}_{i+1} = 0. \tag{6.22}$$

6.2. Tétel. *Legyenek a $\boldsymbol{p}_j$ vektorok a (6.14) és (6.13) egyenlőségben definiált $\boldsymbol{w}_j$ generátorvektorokból számolva, $1 \le j \le i$. Ekkor*

$$\boldsymbol{w}_j^T\boldsymbol{G}\boldsymbol{p}_i = 0, \tag{6.23}$$

ahol $1 \le j \le i-1$. [5, 10. oldal, Lemma 2]

Bizonyítás. A (6.8) egyenlőség és a $\boldsymbol{p}_j$ vektorok konjugáltsága miatt, $\boldsymbol{S}_j^T\boldsymbol{G}\boldsymbol{p}_i = 0$, $1 \le j \le i-1$. A (6.13)-as egyenlőségből, mivel $\boldsymbol{Q}_0 = \boldsymbol{I}$, $\boldsymbol{G}\boldsymbol{p}_i = \boldsymbol{Q}_{j-1}\boldsymbol{G}\boldsymbol{p}_i$, $1 \le j \le$ $\le i-1$. Ezért $\boldsymbol{w}_j^T\boldsymbol{G}\boldsymbol{p}_i = \boldsymbol{w}_j^T\boldsymbol{Q}_{j-1}\boldsymbol{G}\boldsymbol{p}_i$, és a (6.14) egyenlőségből és a konjugáltságból, $\boldsymbol{w}_j^T\boldsymbol{G}\boldsymbol{p}_i = \boldsymbol{p}_j^T\boldsymbol{G}\boldsymbol{p}_i = 0$, $1 \le j \le i-1$. □

Következmény. *Ha a $\boldsymbol{w}_j$ generátorvektorokat a (6.20) egyenlőség szerint választjuk, akkor*

$$\boldsymbol{p}_j^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_i = 0,$$

ahol $0 \le j \le i-2$.

A generátor vektorok fenti választásával, a (6.14) egyenlőség alakja

$$\boldsymbol{p}_{i+1} = \boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_i - \boldsymbol{S}_i(\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{S}_i)^{-1}\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_i\,. \tag{6.24}$$

A (6.24) egyenlőségből, csak az utolsó két tagja nem nulla a $\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{G}\boldsymbol{p}_i$ szorzatnak, ezért a (6.21) egyenlőség az $\boldsymbol{S}_i^T\boldsymbol{G}\boldsymbol{S}_i$ mátrix diagonalitásából azonnal következik [5].

Néhány megjegyzés:

1. A kezdeti $\boldsymbol{p}_1$ vektor tetszőleges. Ha a $\boldsymbol{p}_1$ vektornak a $\boldsymbol{g}_1$ vektort választjuk, akkor az orthodir algoritmust kapjuk [1] [5].

2. Ha $\boldsymbol{p}_1 = \boldsymbol{K}\boldsymbol{g}_1$, akkor a (6.20) egyenlőségből kapott $\{\boldsymbol{x}_i\}$ vektorsorozat megegyezik a (6.16) egyenlőségből kapott $\{\boldsymbol{x}_i\}$ vektorsorozattal, és a stabilitás feltételei hasonlóak [5].

3. Mivel a kezdeti $\boldsymbol{p}_1$ vektor választása tetszőleges, választhatjuk a $\boldsymbol{P}_1$ mátrixnak, tetszőleges számú oszloppal, és így kapjuk a blokk módszereket [5].

4. A 2-tagú módszereknél a konjugált $\boldsymbol{p}_j$ vektorok számítása elválaszthatatlanul összefügg a $\phi(\boldsymbol{x})$ függvény stacionárius pontjának számításával, viszont a 3-

tagú módszereknél a $\boldsymbol{p}_j$ konjugált vektorok generálása teljesen független a $\phi(\boldsymbol{x})$ függvény stacionárius pontjának számításától (módszer 6, 16, 17, 19) [5].

6.2. Szimmetrikus mátrixok

A fő módszerek az (6.1) egyenletrendszer megoldására, ha az $\boldsymbol{A}$ együtthatómátrix szimmetrikus, a következők [5].

Módszer	$\boldsymbol{G}$	$\boldsymbol{h}$	$\boldsymbol{K}$	Név	Referencia
1	$\boldsymbol{A}$	$\boldsymbol{b}$	$\boldsymbol{I}$	cg	[7, 3.5.1-es fejezet], [9]
2	$\boldsymbol{A}^2$	$\boldsymbol{Ab}$	$\boldsymbol{A}^{-1}$	cr	[7, 3.6.1-es fejezet], [9]
3	$\boldsymbol{A}$	$\boldsymbol{b}$	$\boldsymbol{M}^{-1}$	pcg	[1]

Ezek a $\boldsymbol{p}_i$ vektorra 2-tagú rekurrens módszert adnak. Az első kettő, az eredeti konjugált gradiens módszer és a konjugált reziduális módszer b-stabilis és ω-stabilis, ha az $\boldsymbol{A}$ együtthatómátrix pozitív definit. Ha az $\boldsymbol{A}$ együtthatómátrix indefinit, akkor a cg módszer csak $\boldsymbol{b}$-stabilis, a cr módszer pedig csak ω-stabilis. Ezért a módszerek használata indefinit $\boldsymbol{A}$ együtthatómátrix esetén nem ajánlott. A 3-as módszer, a prekondicionált konjugált gradiens módszer b-stabilis és ω-stabilis, ha az $\boldsymbol{A}$ és $\boldsymbol{M}$ mátrixok definitek. Az $\boldsymbol{M}$ prekondicionáló mátrixot úgy választjuk, hogy a $\boldsymbol{KG}$ mátrix kondíciószámát csökkentse, ezzel javítva a módszer konvergenciáját. A fejezet több másik algoritmusát is prekondicionálhatjuk a $\boldsymbol{K}$ mátrix módosításával, de erre külön nem térünk ki [5].

A $\boldsymbol{p}_i$ vektorra 3-tagú rekurrens módszert adnak a következők, ha az $\boldsymbol{A}$ együtthatómátrix szimmetrikus [5].

Módszer	$\boldsymbol{G}$	$\boldsymbol{h}$	$\boldsymbol{K}$	Név	Referencia
4	$\boldsymbol{A}$	$\boldsymbol{b}$	$\boldsymbol{I}$	Nazareth	[10]
5	$\boldsymbol{A}^2$	$\boldsymbol{Ab}$	$\boldsymbol{A}^{-1}$		[5]
6	$\boldsymbol{I}$	$\boldsymbol{A}^{-1}\boldsymbol{b}$	$\boldsymbol{A}$	SYMMLQ	[7, 3.8.1-es fejezet]

A 4-es és 5-ös módszer az 1-es és 2-es módszer 3-tagú változata. A 6-os módszernek nincs 2-tagú megfelelője. Mivel $\boldsymbol{G} = \boldsymbol{I}$, bármilyen ortogonális $\boldsymbol{p}_j$ vektorok megfelelőek lesznek. A probléma az $\boldsymbol{x}_{i+1}$ vektorok kiszámítása, mert a (6.7) egyenlőségben szerepel $\boldsymbol{g}_i = \boldsymbol{x}_1 - \boldsymbol{A}^{-1}\boldsymbol{b}$, és $\boldsymbol{A}^{-1}$ nem ismert. Egy lehetséges megoldás, hogy válasszuk a $\boldsymbol{p}_1$ vektort úgy, hogy $\boldsymbol{p}_1 = \boldsymbol{Ar}_1$, ahol

$$\boldsymbol{r}_i = \boldsymbol{Ax}_i - \boldsymbol{b} \tag{6.25}$$

maradéktag, és indirekt kapjuk a (6.10) egyenlőségben szereplő $\boldsymbol{p}_i^T\boldsymbol{g}_i$-t, de ez numerikusan nemkívánatos [5]. Ha a $\boldsymbol{p}_j$ vektorokat a (6.21) egyenlőség felhasználásával

számoljuk, akkor a (6.8) egyenlőségből, és a $\boldsymbol{G}$ és $\boldsymbol{K}$ mátrixok megfelelő értékeit felhasználva,

$$\boldsymbol{A}\boldsymbol{S}_i = \boldsymbol{S}_{i+1}\boldsymbol{T}_{i+1}\,, \tag{6.26}$$

ahol $\boldsymbol{T}_{i+1}$ az $(i+1)\times i$ méretű bal felső sarokminormátrixa valamilyen tridiagonális mátrixnak [5]. Legyen $\boldsymbol{Q}_{i+1}$ olyan ortogonális mátrix, hogy

$$\boldsymbol{Q}_{i+1}\boldsymbol{T}_{i+1} = \begin{bmatrix} \boldsymbol{U}_i \\ \boldsymbol{0}^T \end{bmatrix}\,, \tag{6.27}$$

ahol $\boldsymbol{U}_i$ felső háromszögmátrix, és $\boldsymbol{0}^T$ sorvektor. Legyen az $\boldsymbol{Y}_i$ mátrix

$$\begin{bmatrix} \boldsymbol{Y}_i & \boldsymbol{y}_{i+1} \end{bmatrix} = \boldsymbol{S}_{i+1}\boldsymbol{Q}_{i+1}^T\,. \tag{6.28}$$

Ekkor a (6.26), (6.27), (6.28) egyenlőségekből

$$\boldsymbol{A}\boldsymbol{S}_i = \boldsymbol{Y}_i\boldsymbol{U}_i\,, \tag{6.29}$$

ahol, mivel az $\boldsymbol{S}_{i+1}$ mátrix oszlopai ortogonálisak, az $\boldsymbol{Y}_i$ mátrix oszlopai szintén ortogonálisak. Ha most minimalizálni akarjuk a $\phi(\boldsymbol{x})$ függvényt az

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{Y}_i\boldsymbol{z} \tag{6.30}$$

hipersík felett, a $\boldsymbol{G}_i$, $\boldsymbol{S}_i$ mátrix és a $\boldsymbol{g}_i$ vektor megfelelő értékeit a (6.7) egyenlőségbe helyettesítve, a (6.11), (6.25) és (6.29) egyenlőségből kapjuk, hogy

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \boldsymbol{Y}_i(\boldsymbol{Y}_i^T\boldsymbol{Y}_i)^{-1}\boldsymbol{U}_i^{-T}\boldsymbol{S}_i^T\boldsymbol{r}_1\,. \tag{6.31}$$

Ez lényegében az SYMMLQ módszere Paige-nek és Saunders-nek [5]. Ők azonban az $\boldsymbol{x}_i$ vektorokat csak segédvektoroknak használták, és az $\boldsymbol{x}_i$ vektorokból számoltak egy közelítő vektorsorozatot az (6.1) egyenletrendszer megoldására, számos más fontos numerikus finomítás mellett. Az algoritmusukat először Fletcher használta mint mimum-hiba algoritmus [5], illetve mivel szükség van a $\boldsymbol{p}_{i+1}$ vektorra az $\boldsymbol{x}_{i+1}$ vektor kiszámításához, tekinthetünk a módszerre mint egy implicit „előretekintő” módszerre [5].

6.3. Nemszimmetrikus mátrixok

Ha az $\boldsymbol{A}$ együtthatómátrix nemszimmetrikus, az 1-6 módszereket nem tudjuk alkalmazni. Az 1-es módszert adja azonban a 7-es módszert és még két másik variációt. [5].

Módszer	$\boldsymbol{G}$	$\boldsymbol{h}$	$\boldsymbol{K}$	Név	Referencia
7	$\boldsymbol{A}^T\boldsymbol{A}$	$\boldsymbol{A}^T\boldsymbol{b}$	$\boldsymbol{I}$	cgne	[7, 3.7.3-as fejezet], [9]
8	$\boldsymbol{I}$	$\boldsymbol{A}^{-1}\boldsymbol{b}$	$\boldsymbol{A}^T\boldsymbol{A}$	Craig's method	[7, 3.7.3-as fejezet]
9	$\boldsymbol{Z}\boldsymbol{A}$	$Z\boldsymbol{b}$	$\boldsymbol{Z}^{-1}$	orthodir	[7, 2.5-ös fejezet]

A 7-es módszer, annak ellenére, hogy b-stabilis és ω-stabilis, általában nem kielégítő, mert a $\boldsymbol{GK}$ mátrix kondíciószáma a négyzete az 1-6 módszerek $\boldsymbol{GK}$ mátrixának kondíciószámának, és emiatt gyakran lassú a konvergencia. A 8-as módszerre hasonlóak igazak, de ennél a módszernél a 6-os módszerhez hasonlóan az $\boldsymbol{A}^{-1}$ mátrixra is szükség van még, az $\boldsymbol{x}_i$ vektorok kiszámításához. Ezen túlléphetünk, ha megszorozzuk balról a (6.16) egyenlőséget az $\boldsymbol{A}^{-T}$ mátrixszal, így nem a $\{\boldsymbol{p}_i\}$ vektorsorozatot, hanem egy $\{\boldsymbol{q}_i\}$ vektorsorozatot generálva, ahol $\boldsymbol{q}_i = \boldsymbol{A}^{-T}\boldsymbol{p}_i$, amikből az $\{\boldsymbol{x}_i\}$ sorozat számolható [5]. A 9-es módszer egy speciális esete az általánosabb Young és Jea módszernek [5], ahol szükséges, hogy $\boldsymbol{Z}$ és $\boldsymbol{ZA}$ mátrixok szimmetrikusak legyenek, ami egy erős megkötés. A stabilitás garantált, ha $\boldsymbol{Z}$ és $\boldsymbol{ZA}$ is definit [5].

Mivel a fenti módszerek egyike se teljesen kielégítő nemszimmetrikus mátrixok esetén, nézzük meg a speciális esetet, amikor a $\boldsymbol{G}$ mátrix alakja

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_{22} \end{bmatrix}, \tag{6.32}$$

vagy

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{0} \end{bmatrix}. \tag{6.33}$$

Innen nem nehéz megmutatni [5], hogy ha a $2n\times 2$-es $\boldsymbol{W}_j$ mátrix alakja

$$\boldsymbol{W}_j = \begin{bmatrix} \boldsymbol{w}_{j1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{w}_{j2} \end{bmatrix}, \qquad j = 1,2,...,i, \tag{6.34}$$

akkor

$$\boldsymbol{Q}_i = \begin{bmatrix} \boldsymbol{Q}_{i1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{i2} \end{bmatrix} \tag{6.35}$$

alakú, és

$$\boldsymbol{P}_{i+1} = \begin{bmatrix} \boldsymbol{u}_{i+1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{v}_{i+1} \end{bmatrix} \tag{6.36}$$

alakú. Így a (6.14) egyenlőségből

$$\boldsymbol{u}_{i+1} = \boldsymbol{Q}_{i1}^T\boldsymbol{w}_{i+1,1} \tag{6.37}$$

[5]. Hasonlóan kaphatjuk a $\boldsymbol{v}_{i+1}$ vektort.

Úgy, mint a (6.14) egyenlőség esetében, a $\boldsymbol{W}_{i+1}$ mátrixot megfelelően kell választani, hogy működő algoritmust kapjunk, és itt is két eset lehetséges [5]. Az első, amikor

$$\boldsymbol{W}_j = \boldsymbol{K}\boldsymbol{F}_j \tag{6.38}$$

alakú, ahol

$$\boldsymbol{F}_j = \begin{bmatrix} \boldsymbol{g}_{j1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{g}_{j2} \end{bmatrix} . \tag{6.39}$$

A $\boldsymbol{g}_{j1}$ és $\boldsymbol{g}_{j2}$ vektorok egyértelmű partíciói a gradiens $\boldsymbol{g}_j$ vektornak a (6.11), (6.32) és (6.33) egyenlőségek miatt. A $\boldsymbol{K}$ mátrix alakja

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}_{22} \end{bmatrix} , \tag{6.40}$$

vagy

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{0} \end{bmatrix} , \tag{6.41}$$

ahol $\boldsymbol{K}_{11}$, $\boldsymbol{K}_{12}$, $\boldsymbol{K}_{21}$, $\boldsymbol{K}_{22}$ konstansok és $\boldsymbol{K} = \boldsymbol{K}^T$. Ebben az esetben meg lehet mutatni [5], hogy a (6.37) egyenlőségből

$$\boldsymbol{u}_{i+1} = \boldsymbol{w}_{i+1,1} - \boldsymbol{u}_i \alpha_{i1} , \tag{6.42}$$

ahol α_{i1}-et úgy választjuk, hogy kielégítse vagy az $\boldsymbol{u}_i^T \boldsymbol{G}_{11} \boldsymbol{u}_{i+1} = 0$ egyenlőséget, ha a $\boldsymbol{G}$ mátrix (6.32) szerint adott, vagy a $\boldsymbol{v}_i^T \boldsymbol{G}_{21} \boldsymbol{u}_{i+1} = 0$ egyenlőséget, ha a $\boldsymbol{G}$ mátrix (6.33) szerint adott [5]. Hasonlóan kaphatjuk a $\boldsymbol{v}_{i+1}$ vektort. A $\boldsymbol{G}$ és a $\boldsymbol{K}$ mátrix lehetséges választásai jelenleg a következők [5].

Módszer	$\boldsymbol{G}$	$\boldsymbol{h}$	$\boldsymbol{K}$	**Név**	**Referencia**
10	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}$	bcg	[7, 3.5.2-es fejezet]
11	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$	Heg	[7, 3.5.4-es fejezet]
12	$\begin{bmatrix} \boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}\boldsymbol{A}^T \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{A}^T\boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{c} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^{-1} \\ \boldsymbol{A}^{-T} & \boldsymbol{0} \end{bmatrix}$	bcr	[5]

Az $\boldsymbol{A}^{-T}$ jelölés jelentése $(\boldsymbol{A}^{-1})^T$. Ezeknél a módszereknél, mint az 1-es és 2-es módszereknél, ha a $\boldsymbol{G}$ és a $\boldsymbol{K}$ mátrix nem definit, a stabilitás nem garantált. A 10-es módszer a Lanczos-féle bikonjugált gradiens módszer, ami se nem b-stabilis, se nem ω-stabilis. A 11-es Hegedüs-féle módszer ω-stabilis, de nem b-stabilis. A 12-es mód-

szer a 2-es módszer általánosítása, és b-stabilis, ha az $\boldsymbol{A}$ együtthatómátrix négyzetes és nemszinguláris [5].

Az algoritmusok következő csoportját úgy kapjuk, hogy a (6.38) egyenlőséget a

$$\boldsymbol{W}_j = \boldsymbol{K}\boldsymbol{G}\boldsymbol{P}_{j-1} \tag{6.43}$$

egyenlőségre cseréljük [5]. Ekkor meg lehet mutatni [5], hogy

$$\boldsymbol{u}_{i+1} = \boldsymbol{w}_{i+1,1} - \boldsymbol{u}_i\alpha_{i1} - \boldsymbol{u}_{i-1}\beta_{i-1,1}\,, \tag{6.44}$$

ahol α_{i1}-et és $\beta_{i-1,1}$-et úgy választjuk, hogy kielégítse az

$$\boldsymbol{u}_{i-1}^T\boldsymbol{G}_{11}\boldsymbol{u}_{i+1} = \boldsymbol{u}_i^T\boldsymbol{G}_{11}\boldsymbol{u}_{i+1} = 0$$

egyenlőséget, ha a $\boldsymbol{G}$ mátrix (6.32) szerint adott, vagy a

$$\boldsymbol{v}_{i-1}^T\boldsymbol{G}_{21}\boldsymbol{u}_{i+1} = \boldsymbol{v}_i^T\boldsymbol{G}_{21}\boldsymbol{u}_{i+1} = 0$$

egyenlőséget, ha a $\boldsymbol{G}$ mátrix (6.33) szerint adott. Úgy, mint a 4-6-os módszereknél, $\boldsymbol{u}_0$ nullvektor és $\boldsymbol{u}_1$ tetszőleges vektor. A (6.44) egyenlőséghez hasonló egyenlőségből kaphatjuk a $\boldsymbol{v}_{i+1}$ vektort. A következők a lehetőségek [5].

Módszer	$\boldsymbol{G}$	$\boldsymbol{h}$	$\boldsymbol{K}$	**Név**	**Referencia**
13	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}$	MRZ	[7, 3.5.3-as fejezet]
14	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$	Heg3	[7, 3.5.3-as fejezet]
15	$\begin{bmatrix} \boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}\boldsymbol{A}^T \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{A}^T\boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{c} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^{-1} \\ \boldsymbol{A}^{-T} & \boldsymbol{0} \end{bmatrix}$	bcr3	[7, 3.6.4-es fejezet]
16	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{A}^{-T}\boldsymbol{c} \\ \boldsymbol{A}^{-1}\boldsymbol{b} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{0} \end{bmatrix}$	QMR	[7, 1.2-es fejezet]
17	$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{A}^{-1}\boldsymbol{b} \\ \boldsymbol{A}^{-T}\boldsymbol{c} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$		[5]

A 13-as módszer a 3-tagú változata a bikonjugált gradiens módszernek. A 14-es és 15-ös módszer a 3-tagú változata a Hegedüs és bikonjugált reziduális módszernek. A 15-ös módszer érdekes tulajdonsága, hogy az $\boldsymbol{A}$ mátrix nem kell, hogy négyzetes legyen, és ha $m\times n$-es n rangú mátrix (tehát $\boldsymbol{A}^T\boldsymbol{A}$ nemszinguláris), akkor a $\boldsymbol{K}_{12}$ mátrixot $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ mátrixnak választva, az algoritmus változatlan és használható. A 14-es módszernek ugyanez a tulajdonsága. A 16-os módszer alkotja az alapját

a QMR módszernek. Csak úgy, mint a SYMMLQ (módszer 6) esetén, a probléma a kvadratikus függvény stacionárius pontjának számolása, mert szükség van $\boldsymbol{A}^{-1}$-re. A 17-es módszereknél ugyanolyan nehézségekbe ütközünk, mint a 6-os és 16-os módszernél, és ezeket hasonlóan is lehet megoldani [5].

A módszerek utolsó csoportja tulajdonképpen az általánosított Lanczos-féle algoritmus [5]. Legyen a (6.21) egyenlőségben a $\boldsymbol{G}$ és $\boldsymbol{K}$ mátrix a (6.32) és a (6.41) egyenlőség szerint adott. Legyen $\boldsymbol{p}_i^T = \begin{bmatrix}\boldsymbol{u}_i^T & \boldsymbol{v}_i^T\end{bmatrix}$. Ha a tetszőlegesen választható $\boldsymbol{p}_1 = \boldsymbol{0}$ és $\boldsymbol{v}_1 = \boldsymbol{0}$, akkor a (6.21) egyenlőségből kivonva, és a konjugáltságot megkövetelve kapjuk, hogy

$$\boldsymbol{u}_{i+1} = \boldsymbol{K}_{12}\boldsymbol{G}_{22}\boldsymbol{v}_i - \boldsymbol{u}_{i-1}\beta_{i-1} \tag{6.45}$$

és $\boldsymbol{v}_{i+1} = \boldsymbol{0}$, minden páros i-re, és

$$\boldsymbol{v}_{i+1} = \boldsymbol{K}_{21}\boldsymbol{G}_{11}\boldsymbol{u}_i - \boldsymbol{v}_{i-1}\delta_{i-1} \tag{6.46}$$

és $\boldsymbol{u}_{i+1} = \boldsymbol{0}$, minden páratlan i-re [5]. Hasonló egyenlőségekhez juthatunk a $\boldsymbol{p}_1 = \boldsymbol{0}$ és az $\boldsymbol{u}_1 = \boldsymbol{0}$ választással [5]. Ez a Golub-Kahan algoritmus általánosítása, és két esetben használható lineáris egyenletrendszerek megoldására [5].

Módszer	$\boldsymbol{G}$	$\boldsymbol{h}$	$\boldsymbol{K}$	**Név**	**Referencia**
18	$\begin{bmatrix}\boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}\boldsymbol{A}^T\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{A}^T\boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{c}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{0} & \boldsymbol{A}^{-1} \\ \boldsymbol{A}^{-T} & \boldsymbol{0}\end{bmatrix}$	bcr3	[7, 3.6.4]
19	$\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{A}^{-1}\boldsymbol{b} \\ \boldsymbol{A}^{-T}\boldsymbol{c}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{0} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0}\end{bmatrix}$	LSQR	[7, 3.8.2]

A 18-as módszer még egy változata a bikonjugált reziduális módszernek, amíg a 19-es módszer a Paige és Saunders féle LSQR algoritmus [5]. Itt is ortogonális transzformációk szükségesek ahhoz, hogy a megoldás következő becslését ki tudjuk számolni, mivel a gradiens számoláshoz szükséges az együtthatómátrix inverze [5].

6.4. Konklúzió

Megmutattuk, hogy nem kevesebb mint tizenkilenc konjugált gradiens algoritmust származtathatunk azzal, hogy $\boldsymbol{G}$ és $\boldsymbol{K}$ két szimmetrikus mátrixot választunk, és a generátorokat két különböző mód egyikével definiáljuk. Az egyik választás két-tagú (Hestenes-Stiefel) rekurrens módszerekhez vezetett, a másik 3-tagú (Lanczos) módszerekhez vezetett. A 2-tagú módszerek stabilitása garantált, ha $\boldsymbol{G}$ és $\boldsymbol{K}$ mátrixok definitek. Broyden a rendszertanába [5] csak 2 vagy 3-tagú rekurrens módszerket vett be. Így fontos módszerek, mint pl. a GCR, GMRES, ORTHOMIN kimaradtak, és pl. az orthodir módszer csak a teljesen szimmetrikus formájában szerepelt.

7. Blokk konjugált gradiens módszer (BlCG)

Jól ismert, hogy a konjugált gradiens módszereket nemszimmetrikus vagy indefinit szimmetrikus együtthatómátrixokra alkalmazva, nullával osztás miatt az algoritmus összeomlása következhet be. Ebben a fejezetben, Broyden [4] cikkére támaszkodva, szükséges és elégséges feltételeket vezetünk le az összeomlás elkerülésére, a Lanczos-féle és a Hestenes-Stiefel-féle blokk konjugált algoritmusok esetén. Ezeket a feltételeket aztán az algoritmusokat meghatározó mátrixok definitségéhez kapcsoljuk. Megmutatjuk, hogy a Lanczos-féle algoritmus stabilisabb, mint a Hestenes-Stiefel-féle algoritmus.

7.1. A Lanczos-féle és a Hestenes-Stiefel-féle algoritmusok

A fejezetben tárgyalt két változata a blokk konjugált gradiens algoritmusnak a

$$\boldsymbol{G}\boldsymbol{X} = \boldsymbol{H} \tag{7.1}$$

egyenletrendszer megoldására alkalmas, ahol $\boldsymbol{G}$ $n\times n$-es valós mátrix, $\boldsymbol{X}$ és $\boldsymbol{H}$ pedig $n\times r$-es valós mátrixok, ahol r általában jóval kisebb, mint n. Ha az $\boldsymbol{X}_i$ mátrix az i-edik becsült megoldása a (7.1) egyenletrendszernek, és ha

$$\boldsymbol{F}_i = \boldsymbol{G}\boldsymbol{X}_i - \boldsymbol{H}\,, \tag{7.2}$$

akkor

$$\boldsymbol{X}_{i+1} = \boldsymbol{X}_i + \boldsymbol{P}_i\boldsymbol{M}_i\,, \tag{7.3}$$

ahol

$$\boldsymbol{M}_i = -\boldsymbol{D}_i^{-1}\boldsymbol{P}_i^T\boldsymbol{F}_i \tag{7.4}$$

és

$$\boldsymbol{D}_i = \boldsymbol{P}_i^T\boldsymbol{G}\boldsymbol{P}_i\,. \tag{7.5}$$

Ekkor az algoritmus leállásának elégséges feltétele, a konjugáltság feltétele, azaz $\boldsymbol{P}_i^T\boldsymbol{G}\boldsymbol{P}_j = \boldsymbol{0}$, ahol $i \neq j$. Ha

$$\overline{\boldsymbol{P}}_i = \left[\begin{array}{cccc} \boldsymbol{P}_1 & \boldsymbol{P}_2 & \ldots & \boldsymbol{P}_i \end{array}\right] \tag{7.6}$$

és

$$\boldsymbol{Q}_i = \boldsymbol{I} - \boldsymbol{G}\overline{\boldsymbol{P}_i}\,\overline{\boldsymbol{D}_i}^{-1}\overline{\boldsymbol{P}_i}^T\,, \tag{7.7}$$

ahol

$$\overline{\boldsymbol{D}_i} = \overline{\boldsymbol{P}_i}^T\boldsymbol{G}\overline{\boldsymbol{P}_i}\,, \tag{7.8}$$

akkor $\boldsymbol{Q}_i\boldsymbol{G}\overline{\boldsymbol{P}_i} = \boldsymbol{0}$, és ha

$$\boldsymbol{P}_{i+1} = \boldsymbol{Q}_i^T\boldsymbol{W}_{i+1}\,, \tag{7.9}$$

akkor $\boldsymbol{P}_{i+1}$ automatikusan kielégíti a konjugáltság feltételét bármilyen $\boldsymbol{W}_{i+1}$ generátormátrixra. A $\overline{\boldsymbol{D}_i} = diag\,(\boldsymbol{D}_j)$ mátrix blokkdiagonális, és feltételezzük, hogy nemszinguláris.

Mivel a $\boldsymbol{Q}_i$ mátrixok nagyok és ritkák, a $\boldsymbol{P}_{i+1}$ mátrixok számítása a (7.9) egyenlőséggel nem praktikus. Ha a $\boldsymbol{W}_i$ generátormátrixokat jól választjuk, akkor jelentős egyszerűsítések lehetségesek. A Lazncos-féle algoritmusnál

$$\boldsymbol{W}_i = \boldsymbol{K}\boldsymbol{G}\boldsymbol{P}_{i-1}\,, \tag{7.10}$$

ahol $\boldsymbol{K}$ egy tetszőleges valós szimmetrikus mátrix, a Hestenes-Stiefel-féle algoritmusnál

$$\boldsymbol{W}_i = \boldsymbol{K}\boldsymbol{F}_i\,. \tag{7.11}$$

Ekkor a (7.9) egyenlőség a Lazncos-féle algoritmusnál

$$\boldsymbol{P}_{i+1} = (\boldsymbol{I} - \boldsymbol{P}_i\boldsymbol{D}_i^{-1}\boldsymbol{P}_i^T\boldsymbol{G} - \boldsymbol{P}_{i-1}\boldsymbol{D}_{i-1}^{-1}\boldsymbol{P}_{i-1}^T\boldsymbol{G})\boldsymbol{K}\boldsymbol{G}\boldsymbol{P}_i\,, \tag{7.12}$$

ahol $\boldsymbol{P}_0 = \boldsymbol{0}$, $\boldsymbol{P}_1$ tetszőleges [4], és

$$\boldsymbol{P}_{i+1} = (\boldsymbol{I} - \boldsymbol{P}_i\boldsymbol{D}_i^{-1}\boldsymbol{P}_i^T\boldsymbol{G})\boldsymbol{K}\boldsymbol{F}_{i+1} \tag{7.13}$$

a Hestenes-Stiefel-féle algortimus esetén, ahol $\boldsymbol{P}_0 = \boldsymbol{0}$ [4].

Mindkét algoritmus stabilis, ha a $\boldsymbol{G}$ és $\boldsymbol{K}$ mátrixok definitek [11]. A Hestenes-Stiefel-féle algoritmus akkor és csak akkor összeomlásmentes, ha a $\boldsymbol{P}_i^T\boldsymbol{G}\boldsymbol{P}_i$ és az $\boldsymbol{F}_i^T\boldsymbol{K}\boldsymbol{F}_i$ mátrixok nemszingulárisak, ami garantált, ha a $\boldsymbol{G}$ és a$\boldsymbol{K}$ mátrixok definitek, és a $\boldsymbol{P}_i$ és az $\boldsymbol{F}_i$ mátrixok teljesrangúak [6]. Broyden a $\boldsymbol{P}_i^T\boldsymbol{G}\boldsymbol{P}_i$ és a $\boldsymbol{P}_i^T\boldsymbol{F}_i$ (ami $\boldsymbol{F}\boldsymbol{K}\boldsymbol{F}_i$ a Hestenes-Stiefel-féle algortimus esetén) mátrixokat vizsgálta, a mögöttes Krylov sorozatok segítségével [4].

7.2. Az összeomlás elkerülésének feltételei

Legyen

$$\overline{\boldsymbol{V}}_i = \left[\begin{array}{cccccc} \boldsymbol{P}_1 & \boldsymbol{K}\boldsymbol{G}\boldsymbol{P}_1 & (\boldsymbol{K}\boldsymbol{G})^2\boldsymbol{P}_1 & \ldots & (\boldsymbol{K}\boldsymbol{G})^{i-1}\boldsymbol{P}_1 \end{array}\right] \tag{7.14}$$

és

$$\boldsymbol{S}_i = \overline{\boldsymbol{V}}_i^T\boldsymbol{G}\overline{\boldsymbol{V}}_i\,. \tag{7.15}$$

Megmutatjuk, hogy a Lanczos-féle blokk konjugált gradiens módszer akkor és csak akkor összeomlásmentes, ha az $\{\boldsymbol{S}_j\}$ mátrixok nemszingulárisak.

7.1. Tétel. *Legyen $\boldsymbol{P}_1$ tetszőleges mátrix, és legyen $j = 1,2,...,i$-re az $\boldsymbol{S}_j$ mátrix nemszinguláris, és számoljuk a $\boldsymbol{P}_j$ mátrixokat a Lanczos algoritmussal (7.5-7.9 egyenlőségek, $\boldsymbol{W}_{i+1} = \boldsymbol{KGP}_i$). Ekkor $j = 1,2,...,i$-re a $\boldsymbol{D}_j$ mátrixok nemszingulárisak, és a $\boldsymbol{D}_{i+1}$ mátrix akkor és csak akkor nemszinguláris, ha az $\boldsymbol{S}_{i+1}$ mátrix nemszinguláris. [4, 45. oldal, Theorem 1]*

Bizonyítás. Lásd [4], 45. oldal. □

Tehát, ha a $\boldsymbol{S}_j$ mátrixok nemszingulárisak, akkor a $\boldsymbol{D}_j$ mátrixok sem azok, és így az algoritmus összemolásmentes. Másrészről, ha az $\boldsymbol{S}_j$ mátrix nemszinguláris $j = = 1,2,....,(i-1)$-re, akkor az $\boldsymbol{S}_i$ mátrix szinguláris, ezért a $\boldsymbol{D}_i$ mátrix szinguláris, és a $\boldsymbol{P}_{i+1}$ és az $\boldsymbol{X}_{i+1}$ mátrixok nem számolhatóak. Ezért az algoritmus összeomlik $\boldsymbol{X}_i$ kiszámolása után. A Lanzos-féle algoritmusnál a összeomlásnak csak ez az egyetlen oka lehet, hogy az $\boldsymbol{S}_i$ mátrix szinguláris [4].

A Hestenes-Stiefel-féle blokk konjugált gradiens módszernél a stabilitást a $\boldsymbol{P}_i^T\boldsymbol{F}_i$ mátrix is befolyásolja [6]. Legyen

$$\boldsymbol{T}_i = \left[\overline{\boldsymbol{V}}_i^T\boldsymbol{F}_1\overline{\boldsymbol{V}}_i^T\boldsymbol{G}\overline{\boldsymbol{V}}_{i-1}\right] . \tag{7.16}$$

7.2. Tétel. *A 7.1-es tétel feltételei álljanak fenn, és ezenfelül legyen az $\boldsymbol{X}_1$ mátrix tetszőleges. Ekkor, $j = 1,2,...,i$-re a $\boldsymbol{P}_j^T\boldsymbol{F}_j$ mátrix akkor és csak akkor nemszinguláris, ha a $\boldsymbol{T}_j$ mátrix nemszinguláris. [4, 47. oldal, Theorem 2]*

Bizonyítás. Lásd [4], 47. oldal . □

A Hestenes-Stiefel-féle blokk konjugált gradiens módszernél az algoritmus esetleges összeomlása nem csak az $\{\boldsymbol{S}_j\}$ mátrixsorozattól, hanem a $\{\boldsymbol{T}_j\}$ mátrixsorozattól is függ [4].

7.3. Tétel. *Legyen $\boldsymbol{X}_1$ tetszőleges mátrix, és legyenek az $\boldsymbol{S}_j$ és a $\boldsymbol{T}_j$ mátrixok nemszingulárisak, $j = 1,2,...,i$, és számoljuk a $\boldsymbol{P}_{j+1}$ mátrixot a Hestenes-Stiefel-féle algoritmussal (7.1-7.9 egyenlőségek, $\boldsymbol{W}_{i+1} = \boldsymbol{KF}_{i+1}$). Ekkor a $\boldsymbol{D}_j$ és a $\boldsymbol{P}_j^T\boldsymbol{F}_j$ mátrixok nemszigulárisak, $j = 1,2,..,i$, és a $\boldsymbol{D}_{i+1}$ mátrix akkor és csak akkor nemszinguláris, ha $\boldsymbol{S}_{i+1}$ nemszinguláris. A $\boldsymbol{P}_{i+1}^T\boldsymbol{F}_{i+1}$ mátrix nemszinguláris, akkor és csak akkor, ha a $\boldsymbol{T}_{i+1}$ mátrix nemszinguláris. [4, 48. oldal, Theorem 3]*

Bizonyítás. Lásd [4], 48. oldal. □

Következmény. *Ha a tétel feltételei teljesülnek, kivéve a $\boldsymbol{T}_i$ mátrix nemszingularitása, akkor a $\boldsymbol{D}_{i+1}$ mátrix szinguláris.*

Ha az $\boldsymbol{S}_j$ mátrix szinguláris, a Lanzos-féle és a Hestenes-Stiefel-féle blokk konjugált gradiens algoritmusok összeomlanak, ha a $\boldsymbol{T}_j$ mátrix szinguláris, akkor csak

a Hestenes-Stiefel-féle algoritmus omlik össze. Ezért a Hestenes-Stiefel-féle blokk konjugált gradiens módszert kevésbé megbízhatónak tekinthetjük [4].

Jelölje HS felső index a Hestenes-Stiefel-féle algoritmust, és L felső index a Lanczos-féle algoritmust. Ha a kétféle módszert ugyanarra a problémára alkalmazzuk, ugyanazokkal a kezdeti értékekkel (a Lanczos-féle módszernél $\boldsymbol{P}_1 = \boldsymbol{K}\boldsymbol{F}_1$), akkor minden j-re $\boldsymbol{Q}_j^{HS} = \boldsymbol{Q}_j^L$ és

$$\boldsymbol{P}_j^{HS} = \boldsymbol{P}_j^L \boldsymbol{L}_{j-1}\,, \tag{7.17}$$

ahol $\boldsymbol{L}_{j-1} = \boldsymbol{M}_1^{HS}\boldsymbol{M}_2^{HS}...\boldsymbol{M}_{j-1}^{HS}$ [4]. A 7.3-as tétel következményéből, ha a $\boldsymbol{T}_{j-1}$ mátrix szinguláris, akkor $\boldsymbol{L}_{j-1}$ mátrix is szinguláris, és a (7.17) egyenlőségből következik, hogy a $\boldsymbol{P}_j^{HS}$ mátrix nem teljesrangú, és így a $\boldsymbol{D}_j^{HS}$ mátrix is szinguláris, akkor is, ha a $\overline{\boldsymbol{V}}_j$ mátrix teljesrangú. Ez egy jelentős különbség a Lanczos-féle módszerhez képest, ahol a $\overline{\boldsymbol{V}}_j$ és a $\overline{\boldsymbol{P}}_j$ mátrixok rangja mindig megegyezik [4].

Legvégül megnézzük az $\{\boldsymbol{X}_i\}$ mátrixsorozatot. A (7.4), (7.5) és (7.17) egyenlőségekből

$$\boldsymbol{M}_j^{HS} - \boldsymbol{L}_{j-1}^{-1}\boldsymbol{M}_j^L\,, \tag{7.18}$$

és így $\boldsymbol{M}_j^L = \boldsymbol{M}_1^{HS}\boldsymbol{M}_2^{HS}...\boldsymbol{M}_j^{HS}$ [4]. Ezt írhatjuk úgy is, hogy $\boldsymbol{M}_j^{HS} = (\boldsymbol{M}_{j-1}^L)^{-1}\boldsymbol{M}_j^L$. A (7.17) és (7.18) egyenlőségből $\boldsymbol{P}_j^{HS}\boldsymbol{M}_j^{HS} = \boldsymbol{P}_j^L\boldsymbol{M}_j^L$, így a két módszerrel generált $\{\boldsymbol{X}_i\}$ mátrixsorozat azonos [4].

7.3. Konklúzió

Ha $\boldsymbol{P}_1 = \boldsymbol{K}\boldsymbol{F}_1$, és a $\boldsymbol{K}$ mátrix nemszinguláris, akkor $\boldsymbol{T}_i = \overline{\boldsymbol{V}}_i^T\boldsymbol{K}^{-1}\overline{\boldsymbol{V}}_i$ [4], és mivel $\boldsymbol{S}_i = \overline{\boldsymbol{V}}_i^T\boldsymbol{G}\overline{\boldsymbol{V}}_i$, láthatjuk, hogy [4]:

1. Ha a $\overline{\boldsymbol{V}}_i$ mátrix nem teljesrangú, akkor az $\boldsymbol{S}_i$ és a $\boldsymbol{T}_i$ mátrixok szinguláris és mindkét algoritmus összeomlik.

2. Ha a $\overline{\boldsymbol{V}}_i$ mátrix teljesrangú és a $\boldsymbol{G}$ mátrix indefinit, akkor mindkét algoritmusnál lehetséges az összeomlás.

3. Ha a $\overline{\boldsymbol{V}}_i$ mátrix teljesrangú és a $\boldsymbol{K}$ mátrix indefinit, akkor a Hestenes-Stiefel-féle algoritmusnál lehetséges az összeomlás.

Az első eset kezelésére léteznek különböző módszerek [4]. A második eset komoly összeomlást eredményezhet, ha i valamilyen értékére az $\boldsymbol{S}_i$ mátrix szinguláris. Ha az $\boldsymbol{S}_i$ mátrix azért szinguláris, mert a $\overline{\boldsymbol{V}}_i$ mátrix nem teljesrangú, akkor az $\boldsymbol{S}_k$ mátrix szinguláris minden $k \geq i$ esetén. Ekkor találhatunk olyan j-lépéses „előretekintő" módszert, amivel megakadályozhatjuk az összeomlást. Ha az $\boldsymbol{S}_i$ mátrix szinguláris, miközben a $\overline{\boldsymbol{V}}_i$ mátrix teljesrangú (mert a $\boldsymbol{G}$ mátrix indefinit), akkor létezik olyan $j \geq 1$, hogy az $\boldsymbol{S}_{i+j}$ mátrix nemszinguláris, vagy az $\boldsymbol{S}_k$ mátrix szinguláris minden

$k \geq i$ esetén, és az összeomlás nem javítható. A harmadik esetnél, ha a $\boldsymbol{G}$ mátrix definit, csak a Hestenes-Stiefel-féle algoritmust használhatjuk.

Nemszimmetrikus vagy indefinit szimmetrikus együtthatómátrixok esetén sokszor megvan a lehetőségünk, hogy olyan algoritmust válasszunk, aminél vagy a $\boldsymbol{G}$, vagy a $\boldsymbol{K}$ mátrix, de nem mindkét mátrix definit. A fenti elemzés alapján ilyenkor érdemes a $\boldsymbol{K}$ mátrixot indefinitnek választani, és a Lanczos-féle algoritmust használni.

8. Matlab tesztfeladatokon való összehasonlítás

A Matlab egy interaktív programcsomag tudományos és mérnöki számítások, szimulációk és grafikus adatmegjelenítés elvégzésére. Előre megírt függvények és eszköztárak segítik a felhasználót a gyors algoritmusfejlesztésben és az adatok grafikus megjelenítésében.

8.1. Tesztspecifikáció

A Matlab `gallery` egy tesztmátrix gyűjtemény. Az

```
[A,B,C,...] = gallery(matname,P1,P2,...)
```

függvényhívás visszaad egy tesztmátrixot vagy tesztmátrixokat, amit a `matname` paraméter specifikál. A `matname` paraméter a mátrixcsalád neve. A használható mátrixcsaládok nevét a

```
help gallery
```

paranccsal, vagy az online Matlab dokumentációban érhetjük el. A `P1,P2,...` paraméterek opcionálisak, és a `matname` által specifikált mátrixcsaládtól függenek. Ezek a paraméterek és a pontos hívási szintaktikák szintén elérhetőek `help gallery` paranccsal, vagy az online Matlab dokumentációban. Ez a tesztmátrix galéria több mint ötvenféle mátrixcsaládot, típust tartalmaz, amelyek hasznosak lehetnek pl. különböző algoritmusok tesztelésénél. Ebben a fejezetben a Matlab `gallery`-vel generálunk teszt együtthatómátrixokat, és ezekkel a teszt együtthatómátrixokkal generálunk lineáris egyenletrendszereket, azaz tesztfeladatokat. A tesztfeladatok segítségével összehasonlítunk néhány, a 6. fejezetben tárgyalt konjugált gradiens módszert.

A teszt együtthatómátrixokkal úgy generálunk tesztfeladatot, hogy az $\boldsymbol{A}$ teszt együtthatómátrixot egy ismert, véletlen generált $\boldsymbol{y}$ vektorral megszorozzuk, és így kapunk egy $\boldsymbol{b}$ vektort. A tesztfeladat az $\boldsymbol{Ax} = \boldsymbol{b}$ lineáris egyenletrendszer, ahol $\boldsymbol{x}$ ismeretlen vektor. Ekkor az $\boldsymbol{y}$ vektor az egyenletrendszer pontos megoldása, azaz

ismerjük a pontos eredményt, az $\boldsymbol{x}$ vektor pedig mindig az egyenletrendszer adott módszerrel számolt megoldása.

A Matlab `gallery`-ből a következő mátrixcsaládokat használjuk.

- dorr
- lehmer
- minij
- moler
- sampling

A tesztfeladatokon a pcg, SYMMLQ, bicgstabl, QMR és az LSQR módszereket hasonlítjuk össze. A módszereket lásd 6. fejezet és [7]. A módszereknél a relatív reziduális toleranciáját 10^{-6} értékre állítjuk, a maximum iteráció számot pedig 1000-re. A módszereknek a saját Matlab implementációját használjuk, a függvények neve megegyezik a módszer nevével. A pcg és SYMMLQ módszereknek az $\boldsymbol{AA}^T$ együtthatómátrixokat adtam át, mivel ezek a módszerek megkövetelik az együtthatómátrix szimmetrikusságát.

A konjugált gradiens módszereken kívül minden tesztfeladatot megoldunk a Matlab `mldivide` függvényével. Az `mldivide` függvényt az `x = A\b` paranccsal is meg lehet hívni. Az `mldivide` megvizsgálja a feladat együtthatómátrixát, és különböző algoritmusok segítségével próbálja a lehető legjobb megoldómódszert kiválasztani. A választás mikéntje tárgyalva van a Matlab dokumentációjában.

8.2. A tesztfeladatok elemzése

A generált tesztmátrixok néhány tulajdonságát foglalja össze a következő táblázat.

Mátrixcsalád	Méret	Kondíciószám	Rang	Pozitív definit
dorr	30×30	$5.49 * 10^5$	30	nem
lehmer	30×30	$8.66 * 10^2$	30	igen
minij	30×30	$1.50 * 10^3$	30	igen
moler	30×30	$3.18 * 10^{17}$	29	igen
sampling	30×30	$2.30 * 10^{16}$	29	nem

A mátrixcsaládokat próbáltam úgy választani, hogy a kapott tesztmátrixok különböző tulajdonságokkal rendelkezzenek.

Esetünkben két együtthatómátrix is rosszul kondicionált, vagy más szóval közel szinguláris, a moler-féle és a sampling-féle együtthatómátrixok. Az $\boldsymbol{A}$ együtthatómátrix rosszul kondicionáltsága azt jelenti, hogy a $\boldsymbol{b}$ vektor nagyon kicsi megváltozása az $\boldsymbol{Ax} = \boldsymbol{b}$ egyenletrendszerben, az $\boldsymbol{x}$ vektor nagyon nagy megváltozását eredményezheti. A kondíciószám „felerősíti" a pontatlanságot. Ha például a $\boldsymbol{b}$ vektor elemei mérési hiba miatt ± 0.01 pontossággal ismertek, akkor az $\boldsymbol{A}$ együtthatómátrix 10^2 nagyságú kondíciószáma azt jelenti, hogy az $\boldsymbol{x}$ megoldásvektor elemei $\pm 100 * 0.01 = \pm 1$ pontosságúak lesznek. Természetesen ezek nem pontos definíciók. A mi tesztfeladataink mentesek bármiféle mérési hibától, mivel ismert megoldásból generáltak. A 10^{16} méretű kondíciószám azonban már a Matlab lebegőpontos számábrázolásából eredő numerikus hibákat is felerősíti. Ez látszik abból is, hogy a Matlab `rank` függvénye szerint a moler-féle és a sampling-féle együtthatómátrixok rangja 29, pedig ez azt jelentené, hogy az együtthatómátrixok szingulárisak, és az egyenletrendszernek nincs megoldása. Valószínűleg a Matlab `rank` függvénye a rosszul kondicionáltság miatt felerősített numerikus hibák miatt hibázik. Itt még talán érdemes megjegyezni, habár a diplomamunka nem használja a determináns fogalmát, hogy egy mátrix szingularitását az alkalmazásokban nem érdemes a determináns segítségével meghatározni, mert a determináns számolásából adódó numerikus pontatlanságok miatt könnyen valótlan következtetésekre juthatunk.

8.3. Teszteredmények

A fő program végzi a teszteket, és tárolja az eredményeket. A fő program forráskódja megtalálható a 11. fejezetben.

A következő táblázatok összefoglalják a különböző módszerek eredményeit. A relatív hiba alatt az

$$\left| \frac{\|\boldsymbol{x}\|_2 - \|\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2} \right|$$

értéket értjük, ahol az $\boldsymbol{y}$ vektor az ismert megoldás, az $\boldsymbol{x}$ vektor pedig az adott módszerrel számolt megoldás. A relatív reziduális

$$\frac{\|\boldsymbol{Ax}_i\|_2}{\|\boldsymbol{Ax}_1\|_2},$$

ahol $\boldsymbol{x}_i$ az i-edik iterációhoz tartozó becslés.

A konvergencia oszlop jelentése	
0	A módszer konvergál az adott toleranciával a maximum iterácó számon belül.
1	A módszer elérte a maximum iterációt de nem konvergált.
2	A prekondicionáló mátrix rosszul kondicionált.
3	A módszer stagnált. (Két egymást követő iteráció eredménye ugyanaz)
4	Valamelyik skalár mennyiség túl kicsi, vagy túl nagy lett a folytatáshoz.

pcg	Konvergencia	Relatív reziduális	Iteráció	Relatív hiba
dorr	1	$9.57 * 10^{-3}$	1000	$8.65 * 10^{-2}$
lehmer	0	$7.60 * 10^{-7}$	30	$6.08 * 10^{-8}$
minij	0	$2.27 * 10^{-7}$	24	$4.99 * 10^{-9}$
moler	0	$3.75 * 10^{-7}$	13	$9.10 * 10^{-3}$
sampling	4	-	0	-

SYMMLQ	Konvergencia	Relatív reziduális	Iteráció	Relatív hiba
dorr	0	$2.78 * 10^{-7}$	52	$4.69 * 10^{3}$
lehmer	0	$4.55 * 10^{-7}$	81	$1.64 * 10^{1}$
minij	0	$8.85 * 10^{-7}$	50	$1.40 * 10^{0}$
moler	0	$1.36 * 10^{-7}$	22	$6.44 * 10^{-1}$
sampling	0	$3.03 * 10^{-7}$	18	$9.71 * 10^{-1}$

bicgstabl	Konvergencia	Relatív reziduális	Iteráció	Relatív hiba
dorr	0	$5.43 * 10^{-9}$	14	$5.43 * 10^{-5}$
lehmer	0	$6.15 * 10^{-7}$	12	$2.67 * 10^{-7}$
minij	0	$9.56 * 10^{-7}$	9	$1.13 * 10^{-5}$
moler	0	$2.50 * 10^{-7}$	5	$9.10 * 10^{-3}$
sampling	1	$2.20 * 10^{-3}$	1000	$6.64 * 10^{8}$

QMR	Konvergencia	Relatív reziduális	Iteráció	Relatív hiba
dorr	0	$9.63 * 10^{-11}$	30	$7.54 * 10^{-11}$
lehmer	0	$6.85 * 10^{-7}$	30	$5.38 * 10^{-8}$
minij	0	$9.94 * 10^{-7}$	22	$4.62 * 10^{-8}$
moler	0	$3.71 * 10^{-7}$	13	$9.10 * 10^{-3}$
sampling	1	$6.64 * 10^{-1}$	1000	$5.06 * 10^{1}$

LSQR	Konvergencia	Relatív reziduális	Iteráció	Relatív hiba
dorr	0	$8.83 * 10^{-7}$	51	$8.00 * 10^{-3}$
lehmer	0	$9.76 * 10^{-7}$	69	$1.40 * 10^{-7}$
minij	0	$7.80 * 10^{-7}$	47	$1.12 * 10^{-9}$
moler	0	$4.15 * 10^{-7}$	21	$9.10 * 10^{-3}$
sampling	0	$4.94 * 10^{-7}$	18	$1.33 * 10^{-5}$

Relatív hiba	pcg	SYMMLQ	bicgstabl	QMR	LSQR
dorr	$8.65 * 10^{-2}$	$4.69 * 10^{3}$	$5.43 * 10^{-5}$	$7.54 * 10^{-11}$	$8.00 * 10^{-3}$
lehmer	$6.08 * 10^{-8}$	$1.64 * 10^{1}$	$2.67 * 10^{-7}$	$5.38 * 10^{-8}$	$1.40 * 10^{-7}$
minij	$4.99 * 10^{-9}$	$1.40 * 10^{0}$	$1.13 * 10^{-5}$	$4.62 * 10^{-8}$	$1.12 * 10^{-9}$
moler	$9.10 * 10^{-3}$	$6.44 * 10^{-1}$	$9.10 * 10^{-3}$	$9.10 * 10^{-3}$	$9.10 * 10^{-3}$
sampling	-	$9.71 * 10^{-1}$	$6.64 * 10^{8}$	$5.06 * 10^{1}$	$1.33 * 10^{-5}$

	mldivide
dorr	$2.54 * 10^{-14}$
lehmer	$3.62 * 10^{-15}$
minij	$2.54 * 10^{-15}$
moler	$8.64 * 10^{0}$
sampling	$5.48 * 10^{-2}$

8.4. Konklúzió

A SYMMLQ módszert és a sampling-féle tesztfeladatot leszámítva összességében „jól teljesítenek" a konjugált gradiens módszerek. A sampling-féle tesztfeladat együtthatómátrixa rosszul kondicionált, kondíció száma 10^{16}, és ráadásul nem pozitív definit.

A pcg módszer a sampling-féle tesztfeladaton összeomlik, de itt az együtthatómátrix nem pozitív definit. Azonban a moler-féle tesztfeladatot $9.10 * 10^{-3}$ relatív hibával megoldja mindösszesen 13 iterációval. A moler-féle tesztfeladat együtthatómátrixa rosszul kondicionált, kondíció száma 10^{17}, de az együtthatómátrixa pozitív definit. A dorr-féle tesztfeladaton a pcg módszer nem éri el a relatív reziduális kívánt toleranciáját, de végül maximum iteráció szám mellett csak $8.65 * 10^{-2}$ lesz az eredmény relatív hibája.

A SYMMLQ módszer érdekes eredménye, hogy az első három „könnyebb" feladatot rosszabb eredménnyel oldja meg mint bármelyik másik konjugált gradiens alapú módszer, de a sampling-féle tesztfeladatot $9.71 * 10^{-1}$ relatív hibával megoldja. A moler-féle tesztfeladatot is $6.44 * 10^{-1}$ relatív hibával megoldja.

A bicgstabl módszer az első négy tesztfeladaton jól teljesít, a moler-féle tesztfeladatot $9.10 * 10^{-3}$ relatív hibával megoldja, ugyanúgy mint a pcg módszer. A nem pozitív definit együtthatómátrixú sampling-féle tesztfeladaton a bicgstabl módszer nem éri el a relatív reziduális kívánt toleranciáját, és nem is tudja megoldani a tesztfeladatot, az eredmény relatív hibája $6.64 * 10^{8}$.

A QMR módszer szintén az első négy tesztfeladatot jól megoldja. A sampling-féle tesztfeladatnál nem éri el a relatív reziduális kívánt toleranciáját, de maximum iteráció szám mellett csak $5.06 * 10^{1}$ lesz az eredmény relatív hibája.

Az LSQR módszer kevesebb mint $8 * 10^{-3}$ relatív hibával mindegyik tesztfeladatot megoldja. Az LSQR módszer pontosabban megoldja a két rosszul kondicionált együtthatómátrixos tesztfeladatot mint az `mldivide`. A sampling-féle tesztfeladatot 10^{-5} nagyságrendű relatív hibával megoldja, ami a legjobb eredmény. Ez két nagyságrenddel kisebb hiba mint az `mldivide` hibája ugyanezen a feladaton.

A mintegy referenciaként betett `mldivide` az első három tesztfeladatot messze a legkisebb hibával oldja meg. Valószínűleg csak a számábrázolás korlátozza. A relatív hibák itt 10^{-14} és 10^{-15} nagyságrendűek. A moler-féle tesztfeladatot azonban pontatlanabbúl oldja meg mint bármelyik konjugált gradiens alapú módszer. A SQMMLQ feladatnál egy nagyságrenddel, a többi konjugált gradiens alapú módszerhez képest 3 nagyságrenddel pontatlanabbul. A sampling-féle tesztfeladatot $5.48 * 10^{-2}$ relatív hibával oldja meg.

9. Összefoglaló

Charles George Broyden (1933. február 3. - 2011. május 20.) angol matematikus, fizikus. Jelentős szerepe volt a kvázi-Newton módszerek kifejlesztésében az 1960-as 70-es években. Az addigi módszerkehez képest jelentősen alacsonyabb számítási igényük miatt, a kávzi-Newton módszerek nagy áttörést jelentettek a nemlineáris optimalizálás területén. Broyden élete második felében a numerikus lineáris algebrára fókuszált, ezen belül is a konjugált gradiens módszerekre és ezek rendszertanára. A diplomamunka főleg a kutatásainak a konjugált gradiens módszerekkel kapcsolatos részét dolgozza fel.

Miért választottam ezt a témát? A lineáris algebra szerepe kiemelkedően fontos rengeteg területen, például jelentős geometriai, fizikai és mérnöki alkalmazásokkal rendelkezik, de a modern társadalomtudományokban is alkalmazzák. Számos más területen is találkozhatunk lineáris algebrával, szinte minden tudományág tartalmaz olyan modelleket, amelyek lineáris egyenletrendszerek megoldására vezetik vissza valamilyen probléma megoldását.

A diplomamunka Broyden rövid életrajzával kezdődik. A diplomamunka első része a további fejezeteket alapozza meg. Ezekben a fejezetekben a további fejezetekhez

szükséges fogalmakat vezettem be, definíciókat adtam meg, tételeket mondtam ki, és tisztáztam a jelöléseket. Szintén az elméleti felvezetés része, a lineáris egyenletrendszerek megoldásának osztályozása. A direkt eljárások részben az LU-felbontást és a Gauss-féle eliminációt mutatom be részletesebben. Az iteratív eljárások utal a későbbiekre.

A további fejezetek Broyden, a numerikus lineáris algebra témakörében elért eredményeit foglalják össze.

A SOR módszer konvergencia kritériumát szimmetrikus együtthatómátrixok esetén Alexander Ostrowski dolgozta ki 1954-ben [12]. Broyden elégséges konvergencia feltételeket adott 1964-ben, szimmetrikus és nemszimmetrikus együtthatómátrixok esetére is [2]. Ezeket az eredményeket dolgozza fel a 5. fejezet.

Az egyik legjobban használható módszer az

$$\boldsymbol{Ax} = \boldsymbol{b}$$

egyenletrendszer megoldására, ahol $\boldsymbol{A}$ valós nemszinguláris ritka mátrix, $\boldsymbol{b}$ pedig valós vektor, a konjugált gradiens módszer és az ebből származtatott különböző módszerek [5]. Hestenes és Siefel eredeti 1952-es módszere [9] csak akkor alkalmazható, ha az $\boldsymbol{A}$ együtthatómátrix szimmetrikus és pozitív definit. Az eredeti módszer óta már rengetek származtatott módszer született, amelyek nem csak szimmetrikus indefinit együtthatómátrixok esetén, de nemszimmetrikus együtthatómátrixok esetén is alkalmazhatóak [5]. A módszerek áttekintése azonban nehézkes lehet, például amiatt, hogy az algoritmusokat a különböző szerzők különböző módokon származtatják. Broyden 1996-os cikkében [5] rendszerezi a konjugált gradiens módszereket. Ezeket az eredményeket foglalja össze a 6. fejezet.

A 7. fejezetben a Lanczos-féle és a Hestenes-Stiefel-féle algortimusokat hasonlítom össze, a blokk konjugált gradiens módszeren keresztül. Broyden az összeomlás elkerüléseinek feltételeit vizsgálta [4].

A 8. fejezetben MATLAB tesztfeladatokon hasonlítom össze a pcg, SYMMLQ, bicgstabl, QMR és az LSQR módszereket. A használt Matlab kód megtalálható a függelékben.

10. Summary

Charles George Broyden (3 February 1933 – 20 May 2011) was an English mathematician, physicist. He had a major part in the development of the quasi-Newton methods in the 1960-70s. The quasi-Newton methods greatly reduced the needed computing capacity to solve nonlinear optimization problems. In the second part of his life Broyden was spending his time on numerical linear algebra. He was de-

aling mainly with the conjugate gradient methods and their taxonomy. This thesis presents his research of the second part of his life.

Why did I chose this topic? Linear algebra is really an important field. It has applications in geometry, physics, engineering, etc., it is even used in modern social sciences. We can say that we can find linear algebra in nearly every field. It is used to model real-life problems with linear systems and to obtain the solution of these systems.

The thesis begins with the short summary of Broyden's life. The first part of the thesis gives the necessary background for the following chapters. The first part of the thesis includes definitions, theorems, and clarifies the notation. It also includes a taxonomy of linear systems. The LU-decomposition and the Gauss-elimination are discussed in more detail.

The remaining chapters summarize the numerical linear algebra results of Broyden.

The convergence criteria of the SOR method for symmetric coefficient matrices were first introduced by Alexander Ostrowski in 1954 [12]. Broyden gave sufficient convergence criteria for symmetric and non-symmetric coefficient matrices in 1964 [2]. This is covered by chapter 5.

One of the best methods for solving the linear system

$$\boldsymbol{Ax} = \boldsymbol{b}$$

where $\boldsymbol{A}$ is large and sparse is the conjugate gradient method or one of its many derivates. The original method, due to Hestenes and Stiefel [9], applies only in the case where $\boldsymbol{A}$ is symmetric and positive definite but many variations have been proposed to deal not only with symmetric indefinite matrices but with non-symmetric matrices as well. However, the overview of these methods can be quite difficult because the different authors use different techniques when deriving the new algorithms. Broyden gives a new taxonomy of the conjugate gradient methods in his publication in 1996 [5]. The result is summarized in chapter 6.

In chapter 7 I compare the Lanczos and Hestenes-Stiefel algorithm through the block conjugate gradient method. Broyden investigated the breakdown conditions of these algorithms [4].

In chapter 8 I compare the pcg, SYMMLQ, bicgstabl, QMR and LSQR method in Matlab. The Matlab source code can be found at the end of the thesis.

11. Függelék

11.1. A fő program

```
number_of_testmatrices = 5;
y = randn(30, 1);

% Setup the linear systems
Test(1).Matrix = 'dorr';
Test(1).A = gallery('dorr', 30);
Test(1).b = Test(1).A*y;
Test(2).Matrix = 'lehmer';
Test(2).A = gallery('lehmer', 30);
Test(2).b = Test(2).A*y;
Test(3).Matrix = 'minij';
Test(3).A = gallery('minij', 30);
Test(3).b = Test(3).A*y;
Test(4).Matrix = 'moler';
Test(4).A = gallery('moler', 30);
Test(4).b = Test(4).A*y;
Test(5).Matrix = 'sampling';
Test(5).A = gallery('sampling', 30);
Test(5).b = Test(5).A*y;

% Save some of the properies of the coefficient matrices
for i = 1 : number_of_testmatrices
    Test(i).cond = cond(Test(i).A);
    Test(i).rank = rank(full(Test(i).A));
    if (issymmetric(Test(i).A));
         Test(i).issymmetric = 'true';
    else
         Test(i).issymmetric = 'false';
    end
    [~,p] = chol(Test(i).A);
    if (p)
        Test(i).posdef = 'false';
    else
        Test(i).posdef = 'true';
    end
end
```

```
for i = 1 : number_of_testmatrices
    % Calculate the result with the pcg method
    [Test(i).pcg.x, Test(i).pcg.flag, Test(i).pcg.relres, ...
    Test(i).pcg.iter]  = ...
    pcg(Test(i).A, Test(i).b, 1e-6, 1000); % use A*A'
    Test(i).pcg.error = abs((norm(y) - norm(Test(i).pcg.x)) ...
    / norm(y));

    % Calculate the result with the SYMMLQ method
    [Test(i).symmlq.x, Test(i).symmlq.flag, ...
    Test(i).symmlq.relres, Test(i).symmlq.iter]  = ...
    symmlq(Test(i).A*Test(i).A', Test(i).b, 1e-6, 1000);
    Test(i).symmlq.error = abs((norm(y) - norm(Test(i).symmlq.x)) ...
    / norm(y));

    % Calculate the result with the bicgstabl method
    [Test(i).bicgstabl.x, Test(i).bicgstabl.flag, ...
    Test(i).bicgstabl.relres, Test(i).bicgstabl.iter]  = ...
    bicgstabl(Test(i).A, Test(i).b, 1e-6, 1000);
    Test(i).bicgstabl.error = abs((norm(y) - ...
    norm(y)) / norm(Test(i).bicgstabl.x));

    % Calculate the result with the QMR method
    [Test(i).qmr.x, Test(i).qmr.flag, Test(i).qmr.relres, ...
    Test(i).qmr.iter]  = ...
    qmr(Test(i).A, Test(i).b, 1e-6, 1000);
    Test(i).qmr.error = abs((norm(y) - norm(Test(i).qmr.x)) ...
    / norm(y));

    % Calculate the result with the LSQR method
    [Test(i).lsqr.x, Test(i).lsqr.flag, Test(i).lsqr.relres, ...
    Test(i).lsqr.iter]  = ...
    lsqr(Test(i).A, Test(i).b, 1e-6, 1000);
    Test(i).lsqr.error = abs((norm(y) - norm(Test(i).lsqr.x)) ...
    / norm(y));

    % Calculate the result using Matlab mldivide
    Test(i).mldivide.x = mldivide(Test(i).A, Test(i).b);
```

```
    Test(i).mldivide.error = abs((norm(y) - norm(Test(i).mldivide.x)) ...
    / norm(y));
end
```

Hivatkozások

[1] S.F. Ashby, T.A. Manteuffel, and P.E. Saylor. A taxonomy for conjugate gradient methods. *SIAM J. Numer. Anal.*, 27:1542–1568, 1990.

[2] C.G. Broyden. On convergence criteria for the method of successive overrelaxation. *Mathematics of Computation*, 18(85):136–141, 1964.

[3] C.G. Broyden. *Basic Matrices.* The Macmillan Press Ltd, 1975.

[4] C.G. Broyden. A breakdown of the block cg method. *Optimization Methods and Software*, 7(1):41–55, 1996.

[5] C.G. Broyden. A new taxonomy of conjugate gradient methods. *Computers Math. Applic.*, 31(4/5):7–17, 1996.

[6] C.G. Broyden and Boschetti M.A. A comparison of three basic conjugate direction methods. *Numerical Linear Algebra with Applications*, 3(6):473–489, 1996.

[7] C.G. Broyden and M.T. Vespucci. *Krylov Solvers for Linear Algebraic Systems*, volume 11 of *Studies in Computational Mathematics.* Elsevier B. V., 2004.

[8] V. Faber and T. Manteuffel. Necessary and sufficient conditions for the existence of a conjugate gradient method. *SIAM J. Numer. Anal.*, 21:352–262, 1984.

[9] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[10] L. Nazareth. A conjugate gradient algorithm without line searches. *Journal of Optimization Theory and Applications*, 23(3):373–387, 1977.

[11] D.P. O'Leary. The block conjugate gradient algorithm and related methods. *Linear Algebra Applics.*, 29:293–322, 1980.

[12] A. Ostrowski. On the linear iteration procedures for symmetric matrices. *Rend. Mat. e Appl. v. 13*, page 140, 1954.

[13] Rózsa Pál. *Lineáris algebra és alkalmazásai.* Tankönyvkiadó, Budapest, 1991.

[14] A. Ralston. *Bevezetés a numerikus analízisbe.* Műszaki Könyvkiadó, Budapest, 1969.