Numerikus módszerek 1
Bizonyítások kidolgozása

Készítette: Kálovits Dorottya

1. Lebegőpontos számok és tulajdonságaik. A Horner-algoritmus.

a) Ismertesse a lebegőpontos számábrázolás modelljét, és definiálja a gépi számokat. Nevezze meg és számítsa ki a számhalmaz nevezetes mennyiségeit (elemszám, M_∞, ε_0). Szemléltesse a halmaz elemeit számegyenesen. Adjon meg két példát a véges szám-ábrázolásból fakadó furcsaságokra.

b) Az input függvény fogalma, tétel az ábrázolt szám hibájáról, ε_1 mennyiség bevezetése és értelmezése.

Definíció: Normalizált lebegőpontos szám

Legyen $m = \sum_{i=1}^{t} m_i \cdot 2^{-i}$, ahol $t \in \mathbb{N}$, $m_1 = 1, m_i \in \{0, 1\}$.
Ekkor az $a = \pm m \cdot 2^k$ $(k \in \mathbb{Z})$ alakú számot *normalizált lebegőpontos számnak* nevezzük.
m: a szám *mantisszája*, hossza t
k: a szám *karakterisztikája*, $k^- \leq k \leq k^+$

Jelölés: $a = \pm[m_1 \dots m_t | k] = \pm 0.m_1 \dots m_t \cdot 2^k$.

Jelölés: $M = M(t, k^-, k^+)$ a gépi számok halmaza, adott $k^-, k^+ \in \mathbb{Z}$ és $t \in \mathbb{N}$ esetén. (Általában $k^- < 0$ és $k^+ > 0$.)

Definíció: Gépi számok halmaza

$$M(t, k^-, k^+) =$$
$$= \left\{ a = \pm 2^k \cdot \sum_{i=1}^{t} m_i \cdot 2^{-i} \; : \; \begin{array}{l} k^- \leq k \leq k^+, \\ m_i \in \{0,1\}, m_1 = 1 \end{array} \right\} \bigcup \{0\}$$

❶ $\frac{1}{2} \leq m < 1$

❷ M szimmetrikus a 0-ra.

❸ M legkisebb pozitív eleme:

$$\varepsilon_0 = [100 \dots 0 | k^-] = \frac{1}{2} \cdot 2^{k^-} = 2^{k^- - 1}$$

❹ M-ben az 1 után következő gépi szám és 1 különbsége:

$$\varepsilon_1 = [100 \dots 01 | 1] - [100 \dots 00 | 1] = 2^{-t} \cdot 2^1 = 2^{1-t}$$

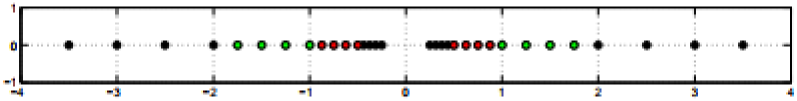
❺ M legnagyobb eleme:

$$M_\infty = [111 \dots 11 | k^+] = 1.00 \dots 00 \cdot 2^{k^+} - 0.00 \dots 01 \cdot 2^{k^+} =$$
$$= (1 - 2^{-t}) \cdot 2^{k^+}$$

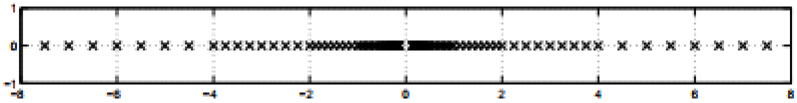
❻ M elemeinek száma (számossága):

$$|M| = 2 \cdot 2^{t-1} \cdot (k^+ - k^- + 1) + 1$$

$$M(3, -1, 2)$$



$$M(4, -2, 3)$$



Mennyi $sin(\pi)$ értéke?

```
1.224646799147353e-016
```

$\mathbb{R}_M := \{x \in \mathbb{R} : \ |x| \le M_\infty\}.$

Definíció: Input függvény

Az $fl : \mathbb{R}_M \to M$ függvényt *input függvénynek* nevezzük, ha

$$fl(x) = \begin{cases} 0 & \text{ha } |x| < \varepsilon_0, \\ \tilde{x} & \text{ha } \varepsilon_0 \le |x| \le M_\infty, \end{cases}$$

ahol $\tilde{x}$ az x-hez legközelebbi gépi szám (a kerekítés szabályai szerint).

Tétel: Input hiba

Minden $x \in \mathbb{R}_M$ esetén

$$|x - fl(x)| \le \begin{cases} \varepsilon_0 & \text{ha } |x| < \varepsilon_0, \\ \frac{1}{2}|x| \cdot \varepsilon_1 & \text{ha } \varepsilon_0 \le |x| \le M_\infty, \end{cases}$$

Következmény: Input hiba

Ha $\varepsilon_0 \le |x| \le M_\infty$, akkor

$$\frac{|x - fl(x)|}{|x|} \le \frac{1}{2} \cdot \varepsilon_1 = 2^{-t}.$$

A hiba tehát lényegében ε_1-től, azaz t-től függ.

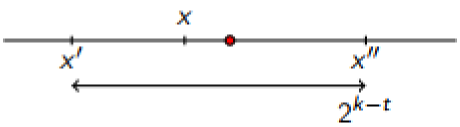
Mennyi a hiba, ha $|x| > M_\infty$?

Bizonyítás:

1. Ha $|x| < \varepsilon_0$, akkor $fl(x) = 0$, így $|x - fl(x)| = |x| < \varepsilon_0$.
2. Ha $|x| \geq \varepsilon_0$ és $x \in M$, akkor $fl(x) = x$, így $|x - fl(x)| = 0$.
3. A meggondolandó eset, amikor $|x| \geq \varepsilon_0$ és $x \notin M$.

Elegendő csak pozitív x-ekkel foglalkoznunk a 0-ra való szimmetria miatt. Keressük meg azt a két szomszédos gépi számot:
$x' < x < x''$ és $x', x'' \in M$, amelyek közrefogják x-et.
Legyen $x' = [1\_ \dots \_|k]$ alakú. Mennyi x' és x'' távolsága?

Ha x-ben az utolsó helyiértékhez 1-et adunk, akkor x''-t kapjuk.
Tehát $x'' - x' = 2^{-t} \cdot 2^k = 2^{k-t}$.



Ha x az intervallum első felében van, akkor $fl(x) = x'$, ha a második felében, akkor $fl(x) = x''$. Ezért x és $fl(x)$ eltérése legfeljebb az intervallum fele, azaz $\frac{1}{2} \cdot 2^k \cdot 2^{-t}$. Vagyis

$$|x - fl(x)| \leq \frac{1}{2} \cdot 2^k \cdot 2^{-t}.$$

Viszont x abszolút értékére, fenti alakját figyelembe véve $0.1 \cdot 2^k = \frac{1}{2} \cdot 2^k \leq |x|$ is teljesül, ezért a becslést így folytathatjuk:

$$|x - fl(x)| \leq |x| \cdot 2^{-t} = \frac{1}{2} \cdot |x| \cdot \underbrace{2^{1-t}}_{\varepsilon_1} = \frac{1}{2} \cdot |x| \cdot \varepsilon_1.$$

$\square$

2. Becslés polinom gyökeinek elhelyezkedésére. A hibaszámítás alapjai.

b) Ismertesse az abszolút és relatív hiba, hibakorlát fogalmát. Mutassa be az alapműveletek hibakorlátaira vonatkozó állításokat, és igazolja a szorzásra vagy osztásra vonatkozó összefüggéseket. Ez alapján mely műveletek elvégzése veszélyes az abszolút és relatív hibára nézve és miért?

c) Igazolja a függvényérték hibakorlátaira vonatkozó tételeket és definiálja függvény adott pontbeli kondíciószámát.

Definíció: Hibák jellemzése

Legyen A egy pontos érték, a pedig egy közelítő értéke. Ekkor:

$\Delta a := A - a$	a közelítő érték (pontos) hibája,
$\lvert\Delta a\rvert := \lvert A - a\rvert$	a közelítő érték abszolút hibája,
$\Delta_a \geq \lvert\Delta a\rvert$	az a egy abszolút hibakorlátja,
$\delta a := \frac{\Delta a}{A} \approx \frac{\Delta a}{a}$	az a relatív hibája,
$\delta_a \geq \lvert\delta a\rvert$	az a egy relatív hibakorlátja.

Tétel: az alapműveletek hibakorlátai

$$\Delta_{a\pm b} = \Delta_a + \Delta_b \qquad \delta_{a\pm b} = \frac{|a|\cdot\delta_a + |b|\cdot\delta_b}{|a\pm b|}$$

$$\Delta_{a\cdot b} = |b|\cdot\Delta_a + |a|\cdot\Delta_b \qquad \delta_{a\cdot b} = \delta_a + \delta_b$$

$$\Delta_{a/b} = \frac{|b|\cdot\Delta_a + |a|\cdot\Delta_b}{b^2} \qquad \delta_{a/b} = \delta_a + \delta_b$$

Megjegyzés: a kapott korlátok két esetben lehetnek nagyságrendileg nagyobbak, mint a kiindulási értékek hibái:

1. $\delta_{a\pm b}$ esetén, amikor közeli számokat vonunk ki egymásból.
2. $\Delta_{a/b}$ esetén, amikor kicsi számmal osztunk.

A szorzás hibája

$$\begin{aligned}\Delta(a\cdot b) &= A\cdot B - a\cdot b = A\cdot B - A\cdot b + A\cdot b - a\cdot b = \\ &= A(B-b) + b(A-a) = A\cdot\Delta b + b\cdot\Delta a = \\ &= (a+\Delta a)\cdot\Delta b + b\cdot\Delta a \approx a\cdot\Delta b + b\cdot\Delta a \\ &(\Delta a\cdot\Delta b \text{ elhanyagolható})\end{aligned}$$

$$|\Delta(a\cdot b)| \leq |a|\cdot|\Delta b| + |b|\cdot|\Delta a| \leq |a|\cdot\Delta_b + |b|\cdot\Delta_a = \Delta_{a\cdot b}$$

A relatív hiba

$$\delta(a\cdot b) = \frac{\Delta(a\cdot b)}{a\cdot b} \approx \frac{a\cdot\Delta b + b\cdot\Delta a}{a\cdot b} = \frac{\Delta b}{b} + \frac{\Delta a}{a} = \delta b + \delta a$$

$$|\delta(a\cdot b)| \leq |\delta a| + |\delta b| \leq \delta_a + \delta_b = \delta_{a\cdot b}$$

Az osztás hibája

$$\begin{aligned}\Delta\left(\frac{a}{b}\right) &= \frac{A}{B}-\frac{a}{b}=\frac{A\cdot b-a\cdot B}{Bb}= \\ &= \frac{A\cdot b-a\cdot b+a\cdot b-a\cdot B}{Bb}=\frac{b\cdot(A-a)-a\cdot(B-b)}{Bb}= \\ &= \frac{b\cdot\Delta a-a\cdot\Delta b}{(b+\Delta b)\cdot b}\approx\frac{b\cdot\Delta a-a\cdot\Delta b}{b^2} \\ &\quad (\Delta b\cdot b \text{ elhanyagolható})\end{aligned}$$

$$\left|\Delta\left(\frac{a}{b}\right)\right|\leq\frac{|b|\cdot|\Delta a|+|a|\cdot|\Delta b|}{b^2}\leq\frac{|b|\cdot\Delta_a+|a|\cdot\Delta_b}{b^2}=\Delta_{a/b}$$

Az osztás relatív hibája

$$\begin{aligned}\delta\left(\frac{a}{b}\right) &= \frac{\Delta(\frac{a}{b})}{\frac{a}{b}}\approx\frac{b\cdot\Delta a-a\cdot\Delta b}{b^2}\cdot\frac{b}{a}= \\ &= \frac{b\cdot\Delta a-a\cdot\Delta b}{b\cdot a}=\frac{\Delta a}{a}-\frac{\Delta b}{b}= \\ &= \delta a-\delta b=\delta\left(\frac{a}{b}\right)\end{aligned}$$

$|\delta\left(\frac{a}{b}\right)|\leq|\delta a|+|\delta b|\leq\delta_a+\delta_b=\delta_{a/b}$ □

1. Tétel: a függvényérték hibája

Ha $f\in C^1(k_{\Delta_a}(a))$ és $k_{\Delta_a}(a)=[a-\Delta_a;a+\Delta_a]$, akkor

$$\Delta_{f(a)}=M_1\cdot\Delta_a,$$

ahol $M_1=\max\{\,|f'(\xi)|\;:\;\xi\in k_{\Delta_a}(a)\,\}$.

Biz.: a Lagrange-féle középértéktétel felhasználásával.

$$\Delta f(a)=f(A)-f(a)=f'(\xi)\cdot(A-a)=f'(\xi)\cdot\Delta a,$$

valamely $\xi\in k_{\Delta_a}(a)$ értékre. Vizsgáljuk az abszolút hibát.

Jó felső becslést adva nyerjük az abszolút hibakorlátot:

$$|\Delta f(a)|=|f'(\xi)|\cdot|\Delta a|\leq M_1\cdot\Delta_a=\Delta_{f(a)},$$

2. Tétel: a függvényérték hibája

Ha $f \in C^2(k_{\Delta_a}(a))$ és $k_{\Delta_a}(a) = [a - \Delta_a; a + \Delta_a]$, akkor

$$\Delta_{f(a)} = |f'(a)|\, \Delta_a + \frac{M_2}{2} \cdot \Delta_a^2,$$

ahol $M_2 = \max\{\, |f''(\xi)| \,:\, \xi \in k_{\Delta_a}(a) \,\}$.

Biz.: a Taylor-formula felhasználásával.

$$\Delta f(a) = f(A) - f(a) = f'(a) \cdot (A - a) + \frac{f''(\xi)}{2} \cdot (A - a)^2,$$

valamely $\xi \in k_{\Delta_a}(a)$ értékre. Vizsgáljuk az abszolút hibát.
Jó felső becslést adva nyerjük az abszolút hibakorlátot:

$$\begin{aligned} |\Delta f(a)| &= |f'(a)| \cdot |\Delta a| + \frac{|f''(\xi)|}{2} \cdot |\Delta a|^2 \leq \\ &\leq |f'(a)| \cdot \Delta_a + \frac{M_2}{2} \cdot \Delta_a^2 = \Delta_{f(a)}, \end{aligned}$$

□

Következmény: függvényérték relatív hibája

Ha Δ_a kicsi, akkor $\delta_{f(a)} = \frac{|a||f'(a)|}{|f(a)|} \cdot \delta_a$.

Definíció: Az f függvény a-beli kondíciószáma

A $c(f, a) = \frac{|a||f'(a)|}{|f(a)|}$ mennyiséget az f függvény a-beli kondíciószámának nevezzük.

3. A Gauss-elimináció és az LU-felbontás algoritmusa.

b) Határozza meg az elimináció és a visszahelyettesítés műveletigényét.

Tétel: A Gauss-elimináció műveletigénye

$$\frac{2}{3}n^3 + \mathcal{O}(n^2)$$

Biz.: Rögzített k-ra: a k. lépés képletéből számolva

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \cdot a_{kj}^{(k-1)} \qquad \begin{aligned} &k = 1, \dots, n-1; \\ &i = k+1, \dots, n; \\ &j = k+1, \dots, n, n+1. \end{aligned}$$

$(n-k)$ osztás, $(n-k)(n-k+1)$ szorzás és $(n-k)(n-k+1)$ összeadás kell.
Összesen $(n-k)(2(n-k)+3)$ művelet. $(n-k =: s)$

$$\sum_{k=1}^{n-1}(n-k)(2(n-k)+3) = \sum_{s=1}^{n-1} s(2s+3) = 2\sum_{s=1}^{n-1} s^2 + 3\sum_{k=1}^{n-1} s =$$
$$= 2\frac{(n-1)n(2n-1)}{6} + 3\frac{(n-1)n}{2} = \frac{2}{3}n^3 + \mathcal{O}(n^2). \quad \square$$

Definíció: $\mathcal{O}(n^2)$ függvény

Az $f(n)$ függvényt $\mathcal{O}(n^2)$-es nagyságrendűnek nevezzük, ha $\frac{f(n)}{n^2}$ korlátos minden $n \in \mathbb{N}$-re.

Tétel: A visszahelyettesítés műveletigénye

$$n^2 + \mathcal{O}(n)$$

Biz.:

$$x_n = \frac{a_{nn+1}^{(n-1)}}{a_{nn}^{(n-1)}}, \quad x_i = \frac{1}{a_{ii}^{(i-1)}}\left(a_{in+1}^{(i-1)} - \sum_{j=i+1}^{n} a_{ij}^{(i-1)} \cdot x_j\right) \quad (i = n-1, \dots, 1).$$

Rögzített i. sorra 1 db osztás, $(n-i)$ szorzás és $(n-i)$ összeadás.
Összesen: $2(n-i)+1$ művelet $(n-i =: s)$.

$$\sum_{s=1}^{n} 2s + 1 = 2 \cdot \frac{n(n+1)}{2} + n = n^2 + \mathcal{O}(n). \quad \square$$

4. A Gauss-elimináció és az LU-felbontás elemzése.

a) Mutassa be a Gauss-elimináció algoritmusát. Adjon szükséges és elégséges feltételeket a GE elakadására illetve végrehajthatóságára. Ismertesse LER megoldását LU-felbontás segítségével. Miért előnyös ennek használata a GE-vel szemben?

b) Ismertesse a részleges és teljes főelemkiválasztás módszereit. Mit mondhatunk az elakadásról részleges főelemkiválasztás alkalmazása esetén? Miért lehet érdemes teljes főelemkiválasztást használni?

Legyen $a_{in+1} := b_i$, azaz $[A|b]$ a tárolási forma.
GE := Gauss-elimináció.

$$A^{(0)} := \left[\begin{array}{cccc|r} a_{11} & a_{12} & \cdots & a_{1n} & a_{1n+1} = b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{2n+1} = b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{nn+1} = b_n \end{array}\right]$$

Célunk: A LER-t egyszerűbb alakra hozni:

1. balról jobbra: a főátló alatt kinullázzuk az elemeket, „előre", GE
2. jobbról balra: a főátló fölött nullázunk, „vissza", visszahelyettesítés

Az 1. egyenletet változatlanul hagyjuk.

Ha $a_{11}^{(0)} \neq 0$, akkor az i-edik egyenletből ($i = 2, 3, \ldots, n$) kivonjuk az 1. egyenlet $\left(\frac{a_{i1}^{(0)}}{a_{11}^{(0)}}\right)$-szeresét: hogy $a_{i1}^{(0)}$ kinullázódjon.
($\rightsquigarrow$ elimináció, kiküszöbölés)

$$A^{(1)} = \left[\begin{array}{ccccc|c} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} & \cdots & a_{1n}^{(0)} & a_{1n+1}^{(0)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} & a_{2n+1}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a_{n2}^{(1)} & a_{n3}^{(1)} & \cdots & a_{nn}^{(1)} & a_{nn+1}^{(1)} \end{array}\right],$$

ahol

$$a_{ij}^{(1)} = a_{ij}^{(0)} - \frac{a_{i1}^{(0)}}{a_{11}^{(0)}} \cdot a_{1j}^{(0)} \qquad (i = 2, \ldots, n;\ j = 2, \ldots, n, n+1).$$

Az 1., 2., $\ldots$, k. egyenleteket változatlanul hagyjuk.

Ha $a_{kk}^{(k-1)} \neq 0$, akkor az i-edik egyenletből ($i = k+1, \ldots, n$) kivonjuk a k-adik egyenlet $\left(\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}\right)$-szeresét: hogy $a_{ik}^{(k-1)}$ kinullázódjon. Ezt a lépést láttuk, amikor a 2. lépésben az 1. lépés eredményét felhasználtuk. Ha 2 helyére k-t írunk, akkor megkapjuk az általános képleteket.

Tétel: A Gauss-elimináció általános lépése

Ha $a_{k,k}^{(k-1)} \neq 0$, akkor a k. lépés képletei

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \cdot a_{kj}^{(k-1)} \qquad \begin{array}{l} k = 1, \ldots, n-1; \\ i = k+1, \ldots, n; \\ j = k+1, \ldots, n, n+1. \end{array}$$

Tétel:

A GE elvégezhető sor és oszlopcsere nélkül
$\Leftrightarrow \quad a_{kk}^{(k-1)} \neq 0 \;\; (k = 1, 2, \ldots, n-1)$.

Biz.: trivi a rekurzióból. $\square$

Definíció: főminorok

Az A főminorai a

$$D_k = \det\left(\begin{bmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{bmatrix}\right), \quad (k = 1, 2, \ldots, n)$$

determinánsok. Ezek az A bal felső $k \times k$-s részmátrixaimak determinánsai.

Tétel:

$D_k \neq 0 \;\; (k = 1, 2, \ldots, n-1) \quad \Leftrightarrow \quad a_{kk}^{k-1} \neq 0 \;\; (k = 1, 2, \ldots, n-1)$.

Biz.: A GE átalakításai determináns tartók, ezért

$$D_k = a_{11} \cdot a_{22}^{(1)} \cdot \ldots \cdot a_{kk}^{(k-1)} = D_{k-1} \cdot a_{kk}^{(k-1)},$$

amiből az állítás adódik. A $D_n \neq 0$ illetve az $a_{nn}^{(n-1)} \neq 0$ feltétel nem szükséges a GE-hoz, csak a LER megoldhatóságához. $\square$

Definíció: LU-felbontás

Az A mátrix LU-felbontásának nevezzük az $L \cdot U$ szorzatot, ha

$$A = LU, \quad L \in \mathcal{L}_1, \quad U \in \mathcal{U}.$$

A Gauss-eliminációt felírhatjuk alsó háromszögmátrixok segítségével:

$$L_{n-1} \cdots L_2 \cdot L_1 \cdot A = U,$$

majd az inverzekkel egyesével átszorozva:

$$A = \underbrace{L_1^{-1} \cdot L_2^{-1} \cdots L_{n-1}^{-1}}_{L} \cdot U = LU.$$

A fenti szorzat is alsó háromszögmátrix. Láttuk az előző tételből, hogy az L mátrix elemeit egy egységmátrixból kapjuk úgy, hogy minden oszlopba ez egyesek alá beletesszük a neki megfelelő ℓ_k vektor nem nulla elemeit (ezek a GE-s hányadosok). Tehát ennek előállításához nem kell több művelet, mint amit a GE-val végzünk.

Tétel: az LU-felbontás „közvetlen" kiszámítása

Az L és U mátrixok elemei a következő képletekkel számolhatók:

$$i \le j \text{ (felső)} \qquad u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik} \cdot u_{kj},$$

$$i > j \text{ (alsó)} \qquad l_{ij} = \frac{1}{u_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} \cdot u_{kj} \right).$$

Ha jó sorrendben számolunk, mindig ismert az egész jobb oldal.

Definíció: részleges főelemkiválasztás

A k-adik lépésben válasszunk egy olyan m indexet, melyre $\left|a_{mk}^{(k-1)}\right|$ maximális ($m \in \{ k, k+1, \ldots, n \}$), majd cseréljük ki a k-adik és m-edik sort.

Definíció: teljes főelemkiválasztás

A k-adik lépésben válasszunk egy olyan (m_1, m_2) indexpárt, melyre $\left|a_{m_1 m_2}^{(k-1)}\right|$ maximális ($m_1, m_2 \in \{ k, k+1, \ldots, n \}$), majd cseréljük ki a k-adik és m_1-edik sort, valamint a k-adik és m_2-edik oszlopot.

Tétel:

A GE elvégezhető sor és oszlopcsere nélkül
$\Leftrightarrow \quad a_{kk}^{(k-1)} \neq 0 \ \ (k = 1, 2, \ldots, n-1).$

Biz.: trivi a rekurzióból. $\square$

Definíció: főminorok

Az A főminorai a

$$D_k = \det\left(\begin{bmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{bmatrix}\right), \quad (k = 1, 2, \ldots, n)$$

determinánsok. Ezek az A bal felső $k \times k$-s részmátrixaimak determinánsai.

Tétel:

$D_k \neq 0 \ \ (k = 1, 2, \ldots, n-1) \quad \Leftrightarrow \quad a_{kk}^{k-1} \neq 0 \ \ (k = 1, 2, \ldots, n-1).$

Biz.: A GE átalakításai determináns tartók, ezért

$$D_k = a_{11} \cdot a_{22}^{(1)} \cdot \ldots \cdot a_{kk}^{(k-1)} = D_{k-1} \cdot a_{kk}^{(k-1)},$$

amiből az állítás adódik. A $D_n \neq 0$ illetve az $a_{nn}^{(n-1)} \neq 0$ feltétel nem szükséges a GE-hoz, csak a LER megoldhatóságához. $\square$

Megj.:

- Numerikus szempontból jobb, ha alkalmazunk főelemkiválasztást. Ezzel a GE-s hányadosaink pontosabbak lesznek.
- Determináns számításakor a cserékkel vigyázni kell!

5. Az LU-felbontás alkalmazása. A Schur-komplementer.

a) Definiálja egy mátrix LU-felbontását. Adjon módszert L és U mátrixok elemenkénti meghatározására, vezesse le az elemekre vonatkozó képleteket. Térjen ki az elemek meghatározásának sorrendjére és a műveletigényre is.

b) Definiálja a Schur-komplementert. Ismertesse a GE megmaradási tételeit (és a kapcsolódó fogalmakat), majd bizonyítsa a determinánsra és szimmetriára vonatkozó pontokat.

Definíció: *LU*-felbontás

Az A mátrix LU-felbontásának nevezzük az $L \cdot U$ szorzatot, ha

$$A = LU, \quad L \in \mathcal{L}_1, \quad U \in \mathcal{U}.$$

A Gauss-eliminációt felírhatjuk alsó háromszögmátrixok segítségével:

$$L_{n-1} \cdots L_2 \cdot L_1 \cdot A = U,$$

majd az inverzekkel egyesével átszorozva:

$$A = \underbrace{L_1^{-1} \cdot L_2^{-1} \cdots L_{n-1}^{-1}}_{L} \cdot U = LU.$$

A fenti szorzat is alsó háromszögmátrix. Láttuk az előző tételből, hogy az L mátrix elemeit egy egységmátrixból kapjuk úgy, hogy minden oszlopba ez egyesek alá beletesszük a neki megfelelő ℓ_k vektor nem nulla elemeit (ezek a GE-s hányadosok). Tehát ennek előállításához nem kell több művelet, mint amit a GE-val végzünk.

Definíció: alsó háromszögmátrix

Az $L \in \mathbb{R}^{n\times n}$ mátrixot *alsó háromszögmátrixnak* nevezzük, ha $i < j$ esetén $l_{ij} = 0$. (A főátló felett csupa nulla.)

$\mathcal{L} := \{\, L \in \mathbb{R}^{n\times n} \;:\; l_{ij} = 0 \; (i < j) \,\}$,
$\mathcal{L}_1 := \{\, L \in \mathbb{R}^{n\times n} \;:\; l_{ij} = 0 \; (i < j), \;\; l_{ii} = 1 \,\}$.

Definíció: felső háromszögmátrix

Az $U \in \mathbb{R}^{n\times n}$ mátrixot *felső háromszögmátrixnak* nevezzük, ha $i > j$ esetén $u_{ij} = 0$. (A főátló alatt csupa nulla.)

$\mathcal{U} := \{\, U \in \mathbb{R}^{n\times n} \;:\; u_{ij} = 0 \; (i > j) \,\}$,
$\mathcal{U}_1 := \{\, U \in \mathbb{R}^{n\times n} \;:\; u_{ij} = 0 \; (i > j), \;\; u_{ii} = 1 \,\}$.

Tegyük fel, hogy

- az $Ax = b$ LER megoldható, és
- rendelkezésünkre áll az $A = LU$ felbontás.

Ekkor $Ax = L \cdot \underbrace{U \cdot x}_{y} = b$ helyett $\left(\frac{2}{3}n^3 + \mathcal{O}(n^2)\right)$

❶ oldjuk meg az $Ly = b$ alsó háromszögű, $\left(n^2 + \mathcal{O}(n)\right)$

❷ majd az $Ux = y$ felső háromszögű LER-t. $\left(n^2 + \mathcal{O}(n)\right)$

Összehasonlításul: egy mátrix-vektor szorzás műveletigénye:
$n \cdot (2n + 1) = 2n^2 + \mathcal{O}(n)$.

Persze valamikor elő kell állítani az LU-felbontást. $\left(\frac{2}{3}n^3 + \mathcal{O}(n^2)\right)$
Előnyös, ha sokszor ugyanaz A: az ILU-algoritmusnál illetve az inverz iterációnál látjuk majd alkalmazását.

Definíció: Schur-komplementer

Tegyük fel, hogy $A_{11} \in \mathbb{R}^{k \times k}$ invertálható mátrix. Az A mátrix A_{11}**-re vonatkozó Schur-komplementere** az

$$[A|A_{11}] := A_{22} - A_{21}A_{11}^{-1}A_{12}$$

$(n-k) \times (n-k)$-s mátrix.

A Schur komplementer azt mutatja, hogy az A_{11}-gyel végzett GE után mely mátrixon kell folytatni az eliminációt. Az új fogalom segítségével könnyebben fogalmazhatjuk meg, hogy a GE mely tulajdonságokat örökíti tovább.

Definíció: szimmetrikus mátrixok

Az A mátrix szimmetrikus, ha $A = A^\top$.

Definíció: pozitív definit mátrixok

Az $A \in \mathbb{R}^{n \times n}$ szimmetrikus mátrix *pozitív definit*, ha

❶ $\langle Ax, x\rangle = x^\top Ax > 0$ bármely $0 \neq x \in \mathbb{R}^n$ esetén; vagy

❷ minden főminorára $D_k = \det(A_k) > 0$; vagy

❸ minden sajátértéke pozitív.

Definíció:

Az A mátrix **szigorúan diagonálisan domináns a soraira**, ha
$|a_{ii}| > \sum_{j=1, j\neq i} |a_{ij}| \quad (i = 1, \dots, n)$.

Definíció:

Az A mátrix **szigorúan diagonálisan domináns az oszlopaira**, ha
$|a_{ii}| > \sum_{j=1, j\neq i} |a_{ji}| \quad (i = 1, \dots, n)$.

Definíció:

Az A mátrix **fél sávszélessége** $s \in \mathbb{N}$, ha

$$\begin{aligned} \forall\ i,j : |i-j| > s : \ \ a_{ij} = 0 \ \text{ és} \\ \exists\ k,l : |k-l| = s : a_{kl} \neq 0. \end{aligned}$$

Definíció:

Az A mátrix **profilja** sorokra a $(k_1, \ldots, k_n)$, oszlopokra az $(l_1, \ldots, l_n)$ szám n-sek, melyekre

$$\begin{aligned} \forall\ j = 1, \ldots, k_i : \ \ a_{ij} = 0 \ \text{ és } \ a_{i,k_i+1} \neq 0, \\ \forall\ i = 1, \ldots, l_j : \ \ a_{ij} = 0 \ \text{ és } \ a_{l_j+1,j} \neq 0. \end{aligned}$$

Soronként és oszloponként az első nem nulla elemig a nullák száma.

Tétel: megmaradási tételek a GE-ra

A GE során a következő tulajdonságok öröklődnek A-ról a Schur-komplementerre:

1. $\det(A) \neq 0 \quad \Rightarrow \quad \det([A|A_{11}]) \neq 0$
2. A szimmetrikus $\Rightarrow$ $[A|A_{11}]$ szimmetrikus
3. A pozitív definit $\Rightarrow$ $[A|A_{11}]$ pozitív definit
4. A szig. diag. dom. $\Rightarrow$ $[A|A_{11}]$ szig. diag. dom.
5. $[A|A_{11}]$ fél sávszélessége $\leq$ A fél sávszélessége
6. A GE során a profilnál a soronkénti és oszloponkénti nullák az első nem nulla elemig megmaradnak.

Biz.: 1.) Determináns:
Mivel a GE determináns tartó, így $\det(A) = \det(A^{(1)}) \neq 0$.

$$A^{(1)} = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & [A|A_{11}] \end{array}\right]$$

$$0 \neq \det(A^{(1)}) = \underbrace{\det(A_{11})}_{\neq 0} \cdot \det([A|A_{11}]) \quad \Leftrightarrow \quad \det([A|A_{11}]) \neq 0$$

$\square$

2.) Szimmetria:
Ha A szimmetrikus, akkor A_{11} és A_{22} is az, továbbá $A_{21}^{\top} = A_{12}$.

$$\begin{aligned} [A|A_{11}]^{\top} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{\top} = A_{22}^{\top} - A_{12}^{\top}(A_{11}^{-1})^{\top}A_{21}^{\top} = \\ &= A_{22}^{\top} - A_{12}^{\top}(A_{11}^{\top})^{-1}A_{21}^{\top} = A_{22} - A_{21}A_{11}^{-1}A_{12} = [A|A_{11}] \end{aligned}$$

$\square$

6. A Cholesky-féle felbontás.

a) Az LDU-felbontás fogalma, előállítása. Szimmetrikus mátrix felbontására vonatkozó tétel.

Definíció: *LDU*-felbontás

Az $A \in \mathbb{R}^{n\times n}$ mátrix LDU-felbontásának nevezzük az $A = L \cdot D \cdot U$ szorzatot, ha $L \in \mathcal{L}_1$ alsó háromszögmátrix, D diagonális mátrix és $U \in \mathcal{U}_1$ felső háromszögmátrix.

Előállítás LU**-felbontásból:**

Az $A = L \cdot \widetilde{U}$ felbontásban $L \in \mathcal{L}_1$ jó, $D = \operatorname{diag}(\widetilde{u}_{11}, \ldots, \widetilde{u}_{nn})$.
A keresett $U \in \mathcal{U}_1$ mátrixot úgy kapjuk, hogy $U = D^{-1}\widetilde{U}$, azaz minden i-re $\widetilde{U}$ i. sorát $\widetilde{u}_{ii}$-vel osztjuk. Ekkor

$$A = L\widetilde{U} = LD \cdot \underbrace{(D^{-1}\widetilde{U})}_{U} = LDU.$$

Tétel: az LDU-felbontás „közvetlen" kiszámítása

Az L, D és U mátrixok eleme:t jó sorrendben (lásd LU-felbontás) számolva a jobboldalon mindig ismert értékek lesznek:

$$i < j \text{ (felső)} \qquad u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik} \cdot d_{kk} \cdot u_{kj},$$

$$i = j \text{ (diag)} \qquad d_{ii} = a_{ii} - \sum_{k=1}^{i-1} l_{ik} \cdot d_{kk} \cdot u_{ki},$$

$$i > j \text{ (alsó)} \qquad l_{ij} = \frac{1}{d_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} \cdot d_{kk} \cdot u_{kj} \right).$$

A képleteket az $A = L\widetilde{U}$ felbontás „közvetlen" képleteiből kapjuk:

$$\widetilde{u}_{ii} \;\mapsto\; d_{ii}, \quad \widetilde{u}_{kj} \;\mapsto\; d_{kk}u_{kj}.$$

Tétel: Szimmetrikus mátrix LDU-felbontása

Ha A szimmetrikus mátrix, akkor az LDU-felbontásában $U = L^\top$.

Biz.: az $A = LDU$ felbontás bal oldalát szorozzuk L^{-1}-zel, jobb oldalát $(L^{-1})^\top$-tal:

$$L^{-1}A(L^{-1})^\top = L^{-1} \cdot (LDU) \cdot (L^{-1})^\top = DU(L^{-1})^\top.$$

A bal oldali mátrixról tudjuk, hogy szimmetrikus, a jobboldali felső háromszögmátrix. Ebből következik, hogy a jobboldali mátrix diagonális mátrix. $U(L^{-1})^\top \in \mathcal{U}_1$, így $U(L^{-1})^\top = I$.

$$U(L^{-1})^\top = I \quad \Leftrightarrow \quad U(L^\top)^{-1} = I \quad \Leftrightarrow \quad U = L^\top$$

$\square$

Következmény:

- Szimmetrikus mátrix esetén az LDU-felbontás megtartja a szimmetriát. A teljes mátrix helyett elég pl. az alsó háromszög részét tárolni. Az $A = LDU$ felbontás valójában $LDL^\top$-felbontás lesz, ahol szintén elég L, D-t tárolni. Ezzel a tárolás- és műveletigény kb. a felére csökken $(\frac{1}{3}n^3 + \mathcal{O}(n^2))$.
- Szimmetrikus mátrix esetén az $LDL^\top$-felbontás GE-val közvetlenül is elkészíthető.

7. A QR-felbontás.

a) Definiálja a QR-felbontást és vezesse le az előállítására alkalmas Gramm–Schmidt ortogonalizációs eljárást. Milyen feltétel garantálja, hogy az algoritmus nem akad el?

b) Mutassa be az ortogonalizációs eljárás normálás nélküli változatát, és az utólagos normálás módját. Hogyan alkalmazható a QR-felbontás LER megoldására? Vesse össze az LU-felbontáson alapuló megoldással (műveletigény, alkalmazhatóság).

Definíció: *QR*-felbontás

Az $A \in \mathbb{R}^{n\times n}$ mátrix QR-felbontásának nevezzük a $Q \cdot R$ szorzatot, ha $A = QR$, ahol $Q \in \mathbb{R}^{n\times n}$ ortogonális mátrix, $R \in \mathcal{U}$ pedig felső háromszögmátrix.

Tétel: *QR*-felbontás létezése és egyértelműsége

Ha $\det A \neq 0$, (vagyis az A oszlopvektorai lineárisan függetlenek), akkor A-nak létezik QR-felbontása.
Ha még feltesszük, hogy $r_{ii} > 0 \ \forall\, i$-re, akkor egyértelmű is.

Biz.: Létezés: A bizonyítást a Gram–Schmidt-féle ortogonalizációs eljárás adja: az A mátrix oszlopaiból – amelyek a feltétel értelmében lineárisan függetlenek – előállítjuk a Q oszlopait és R ismeretlen elemeit.

Definíció: Gram–Schmidt-ortogonalizáció (normálás nélkül)

Adott az $a_1, \ldots, a_n \in \mathbb{R}^n$ lineárisan független vektorrendszer.

1. $\widetilde{q_1} := a_1,$
2. $\widetilde{r_{11}} := 1$

   A k-adik lépésben ($k = 2, \ldots, n$):
3. $\widetilde{r_{jk}} := \dfrac{\langle a_k, \tilde{q}_j\rangle}{\langle \tilde{q}_j, \tilde{q}_j\rangle} \quad (j = 1, \ldots, k-1),$
4. $\widetilde{q_k} := a_k - \sum\limits_{j=1}^{k-1} \widetilde{r_{jk}} \cdot \tilde{q}_j,$
5. $\widetilde{r_{kk}} := 1 \quad$ (nem normálunk),

Az így nyert $\widetilde{q_1}, \ldots, \widetilde{q_n} \in \mathbb{R}^n$ vektorrendszer ortogonális.

Megj.: Levezetése teljesen hasonló. Kézi számolásra alkalmasabb. Ne felejtsünk el normálni...

Tegyük fel, hogy

- az $Ax = b$ LER megoldható, és
- rendelkezésünkre áll az $A = QR$ felbontás.

Ekkor $Ax = Q \cdot \underbrace{R \cdot x}_{y} = b$ helyett $\left(\frac{2}{3}n^3 + \mathcal{O}(n^2)\right)$

❶ a $Qy = b$ LER megoldása: $y = Q^\top b$, $(2n^2 + \mathcal{O}(n))$

❷ az $Rx = y$ LER-t oldjuk meg. $(n^2 + \mathcal{O}(n))$

Együtt is írható: oldjuk meg az $Rx = Q^\top b$ LER-t.

Persze valamikor elő kell állítani a QR-felbontást. $(2n^3 + \mathcal{O}(n^2))$
Előnyös, ha sokszor ugyanaz A, lásd QR-algoritmus (Num. mód. 2A). Így numerikusan stabilabb a LER megoldása.

Definíció: ortogonális mátrix

Egy $Q \in \mathbb{R}^{n\times n}$ mátrix *ortogonális*, ha az inverze a transzponáltja, azaz

$$Q^\top Q = I.$$

Megj.: Ekkor $QQ^\top = I$ is teljesül. $(Q^{-1} = Q^\top)$

Definíció: skaláris szorzat

Az $x, y \in \mathbb{R}^n$ vektorok *skaláris szorzata*

$$\langle x, y\rangle := y^\top x = \sum_{k=1}^{n} x_k \cdot y_k.$$

Definíció: ortonormált rendszer

A $q_1, \ldots, q_n \in \mathbb{R}^n$ vektorok *ortonormált rendszert* alkotnak, ha

$$\langle q_i, q_j\rangle = \begin{cases} 0 & \text{ha } i \neq j, \\ 1 & \text{ha } i = j. \end{cases}$$

Állítás: ortogonális mátrixok oszlopvektorairól

A $Q \in \mathbb{R}^{n\times n}$ ortogonális mátrix oszlopai, mint vektorok ortonormált rendszert alkotnak.

Biz.: Gondoljunk bele: $Q^\top Q = I$. □

Definíció: ortogonális rendszer

A $q_1, \ldots, q_n \in \mathbb{R}^n$ vektorok *ortogonális rendszert* alkotnak, ha

$$\langle q_i, q_j \rangle = 0 \qquad (i \neq j).$$

Állítás: ortogonális rendszerekből álló mátrixokról

Ha a $q_1, \ldots, q_n \in \mathbb{R}^n$ vektorok ortogonális rendszert alkotnak, akkor a $Q := (q_1, \ldots, q_n) \in \mathbb{R}^{n\times n}$ mátrix esetén a $Q^\top Q$ szorzatmátrix diagonális. ($QQ^\top$ általában nem.)

Biz.: Gondoljunk bele: $Q^\top Q = D$ diagonális mátrix. □

Állítás: ortogonális mátrixok szorzata

Ha $Q_1, Q_2 \in \mathbb{R}^{n\times n}$ ortogonális mátrixok, akkor a szorzatuk, $Q_1 Q_2$ is ortogonális.

Biz.: Tudjuk, hogy $Q_1^\top Q_1 = I$ és $Q_2^\top Q_2 = I$.

Kell, hogy $Q_1 Q_2$ is ortogonális.

Vizsgáljuk:

$$(Q_1 Q_2)^\top (Q_1 Q_2) = Q_2^\top \underbrace{Q_1^\top Q_1}_{I} Q_2 = Q_2^\top Q_2 = I.$$

Definíció: QR-felbontás

Az $A \in \mathbb{R}^{n\times n}$ mátrix QR-felbontásának nevezzük a $Q \cdot R$ szorzatot, ha $A = QR$, ahol $Q \in \mathbb{R}^{n\times n}$ ortogonális mátrix, $R \in \mathcal{U}$ pedig felső háromszögmátrix.

Definíció: Gram–Schmidt-féle ortogonalizáció

Adott az $a_1, \ldots, a_n \in \mathbb{R}^n$ lineárisan független vektorrendszer.

1. $r_{11} := \|a_1\|_2$,
2. $q_1 := \dfrac{1}{r_{11}} a_1$ („lenormáljuk”).
   A k-adik lépésben ($k = 2, \ldots, n$):
3. $r_{jk} := \langle a_k, q_j \rangle \quad (j = 1, \ldots, k-1)$,
4. $s_k := a_k - \sum_{j=1}^{k-1} r_{jk} \cdot q_j$,
5. $r_{kk} := \|s_k\|_2$ (s_k segédvektor hossza),
6. $q_k := \dfrac{1}{r_{kk}} s_k$ („lenormáljuk”).

Az így nyert $q_1, \ldots, q_n \in \mathbb{R}^n$ vektorrendszer ortonormált.

Normálás utólag:

- $A = \widetilde{Q}\widetilde{R}$,
- $D := \widetilde{Q}^\top \widetilde{Q}$, azaz $D = \operatorname{diag}\left(\langle q_1, q_1\rangle, \dots \langle q_n, q_n\rangle\right)$,
- $A = \underbrace{\widetilde{Q} \cdot \sqrt{D}^{-1}}_{Q} \cdot \underbrace{\sqrt{D} \cdot \widetilde{R}}_{R} = Q \cdot R$,

  azaz $\widetilde{Q}$ oszlopait, mint vektorokat leosztjuk azok hosszával (normáljuk őket), $\widetilde{R}$ sorait pedig szorozzuk ugyanezekkel az értékekkel.
- Közvetlenül a $\sqrt{D} = \operatorname{diag}\left(\|q_1\|_2, \dots, \|q_n\|_2\right)$ alakkal is dolgozhatunk.

Tétel: A Gram–Schmidt-ortogonalizáció műveletigénye

A szorzások és osztások száma

$$2n^3 + \mathcal{O}(n^2),$$

valamint n darab négyzetgyökvonás is szükséges.

Tétel: Az LU-felbontás műveletigénye

$$\frac{2}{3}n^3 + \mathcal{O}(n^2)$$

Tétel: LU-felbontás létezése és egyértelműsége (főminorokkal)

- Ha $D_k \neq 0 \ \ (k = 1, \dots, n-1)$, akkor létezik az A mátrix LU-felbontása és $u_{kk} \neq 0 \ \ (k = 1, \dots, n-1)$.
- Ha $\det(A) \neq 0$, akkor a felbontás egyértelmű.

Tétel: QR-felbontás létezése és egyértelműsége

Ha $\det A \neq 0$, (vagyis az A oszlopvektorai lineárisan függetlenek), akkor A-nak létezik QR-felbontása.
Ha még feltesszük, hogy $r_{ii} > 0 \ \forall i$-re, akkor egyértelmű is.

8. A Householder-transzformáció.

a) Definiálja a Householder-transzformációt, ismertesse geometriai tartalmát, vezesse le elemi tulajdonságait. Mutassa be a transzformáció alkalmazásának módját vektorra illetve mátrixra (mindkét irányból), adja meg e számítások műveletigényeit.

Definíció: vektorok „hossza"

Az $\mathbb{R}^n$-beli v vektorok hagyományos értelemben vett hosszát, avagy „kettes normáját" jelölje $\|.\|_2$.
A következőképpen számolható:

$$\|v\|_2 := \sqrt{\langle v, v\rangle} = \sqrt{v^\top v} = \left(\sum_{k=1}^{n} v_i^2\right)^{\frac{1}{2}}$$

Definíció: Householder-mátrix

A $H = H(v) \in \mathbb{R}^{n\times n}$ mátrixot *Householder-mátrixnak* nevezzük, ha

$$H(v) = I - 2vv^\top,$$

ahol $v \in \mathbb{R}^n$ és $\|v\|_2 = 1$.

Tétel: tetszőleges tükrözés Householder-mátrixszal

Legyen $a, b \in \mathbb{R}^n$, $a \neq b$ és $\|a\|_2 = \|b\|_2 \neq 0$. Ekkor a

$$v = \pm\frac{a-b}{\|a-b\|_2} \text{ választással } H(v)\cdot a = b.$$

Biz.: Ismerve, hogy $H(v) = I - 2vv^\top$, számoljuk végig a $H(v)\cdot a$ szorzatot. Közben használjuk ki, hogy $\|a\|_2 = \|b\|_2$, azaz $a^\top a = b^\top b$, valamint a skaláris szorzás kommutatív, azaz $a^\top b = b^\top a$.

$$\left(I - 2\frac{(a-b)(a-b)^\top}{\|a-b\|_2^2}\right)\cdot a = a - \frac{2(a-b)(a^\top a - b^\top a)}{(a-b)^\top(a-b)} =$$

$$= a - \frac{2(a-b)(a^\top a - b^\top a)}{a^\top a - a^\top b - b^\top a + b^\top b} = a - \frac{2(a-b)(a^\top a - b^\top a)}{2(a^\top a - b^\top a)} =$$

$$= a - (a-b) = b.$$

Tehát valóban, két különböző, de azonos hosszúságú vektor átvihető egymásba egy Householder-transzformáció által. $\square$

Megjegyzés: Egyébként $H(v)\cdot b = a$ is teljesül.

Példa: Householder-féle tükrözés

Határozzuk meg azt a Householder-féle transzformációt, amely a következő a vektort $b = k \cdot e_1$ alakúra hozza. Ellenőrzésképpen végezzük is el a transzformációt.

$$a = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

A jó előjel választás σ-nak -1, mert a első eleme pozitív.

$$\sigma = -\|a\|_2 = -\sqrt{2^2 + (-2)^2 + 1^2} = -3$$

Ezzel az előjel választással stabilabb lesz az osztásunk v előállításban.

$$a - \sigma e_1 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - (-3) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$$

Látjuk, hogy valójában egyetlen műveletet kellett elvégeznünk a vektor első elemén. Ezzel a σ előjelválasztással elérjük, hogy $\|a - \sigma e_1\|_2 \geq \|a\|_2$.

$$\|a - \sigma e_1\|_2 = \sqrt{5^2 + (-2)^2 + 1^2} = \sqrt{30}$$

$$v = \frac{a - \sigma e_1}{\|a - \sigma e_1\|_2} = \frac{1}{\sqrt{30}} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} \text{ jó választás.}$$

Ellenőrizzük végezzük el a transzformációt a-n:
$H(v) \cdot a = a - 2v \underbrace{(v^\top a)}_{\in \mathbb{R}} = a - 2(v^\top a)v.$

$$H(v) \cdot a = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - 2 \cdot \underbrace{\frac{1}{\sqrt{30}} \begin{bmatrix} 5 & -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}}_{15} \cdot \frac{1}{\sqrt{30}} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix} = \sigma \cdot e_1 \quad \checkmark$$

$\square$

9. Mátrixnormák és tulajdonságaik I.

(a) Definiálja a vektornormát, mátrixnormát és indukált normát, mutassa meg, hogy utóbbi mindig mátrixnorma. Adjon meg példákat is. Igazolja mátrix tetszőleges normája és a spektrálsugara közti egyenlőtlenséget.

Definíció: vektornorma

Legyen $n \in \mathbb{N}$ rögzített. Az $\|.\| : \mathbb{R}^n \to \mathbb{R}$ leképezést vektornormának nevezzük, ha:

1. $\|x\| \geq 0 \quad (\forall x \in \mathbb{R}^n)$,
2. $\|x\| = 0 \iff x = 0$,
3. $\|\lambda \cdot x\| = |\lambda| \cdot \|x\| \quad (\forall \lambda \in \mathbb{R},\ \forall x \in \mathbb{R}^n)$,
4. $\|x + y\| \leq \|x\| + \|y\| \quad (\forall x, y \in \mathbb{R}^n)$.

Azaz a leképezés „pozitív", „pozitív homogén" és „szubadditív" (háromszög-egyenlőtlenség). Ezek a vektornormák *axiómái*.

Állítás: Gyakori vektornormák $(1, 2, \infty)$

A következő formulák vektornormákat **definiálnak** $\mathbb{R}^n$ felett:

- $\|x\|_1 := \sum_{i=1}^{n} |x_i|$ (Manhattan-norma),
- $\|x\|_2 := \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$ (Euklideszi-norma),
- $\|x\|_\infty := \max_{i=1}^{n} |x_i|$ (Csebisev-norma).

Definíció: mátrixnorma

Legyen $n \in \mathbb{N}$ rögzített. Az $\|.\| : \mathbb{R}^{n \times n} \to \mathbb{R}$ leképezést mátrixnormának nevezzük, ha:

1. $\|A\| \geq 0 \quad (\forall A \in \mathbb{R}^{n \times n})$,
2. $\|A\| = 0 \iff A = 0$,
3. $\|\lambda \cdot A\| = |\lambda| \cdot \|A\| \quad (\forall \lambda \in \mathbb{R},\ \forall A \in \mathbb{R}^{n \times n})$,
4. $\|A + B\| \leq \|A\| + \|B\| \quad (\forall A, B \in \mathbb{R}^{n \times n})$,
5. $\|A \cdot B\| \leq \|A\| \cdot \|B\| \quad (\forall A, B \in \mathbb{R}^{n \times n})$.

Ugyanaz, mint a vektornormáknál, plusz: „szubmultiplikativitás".
Ezek a mátrixnormák axiómái.

Definíció: indukált norma, természetes mátrixnormák

Legyen $\|.\|_v : \mathbb{R}^n \to \mathbb{R}$ tetszőleges vektornorma. Ekkor a

$$\|.\| : \mathbb{R}^{n\times n} \to \mathbb{R}, \qquad \|A\| := \sup_{x\neq 0} \frac{\|Ax\|_v}{\|x\|_v}$$

függvényt a $\|.\|_v$ *vektornorma által indukált mátrixnormának* hívjuk. Egy mátrixnormát *természetesnek* nevezünk, ha van olyan vektornorma, ami indukálja.

Tétel: indukált normák

Az „indukált mátrixnormák" valóban mátrixnormák.

Biz.: Be kell látni, hogy a megadott alak teljesíti a mátrixnorma axiómáit.

1. Az $\|A\|$ értéke nemnegatív, hiszen vektorok normájának (nemnegatív számok) hányadosainak szuprémuma.
2. Ha $A = 0$, azaz nullmátrix, akkor $\|Ax\|_v = 0$ minden x vektorra, így a szuprémum értéke is 0. Valamint megfordítva, ha a szuprémum 0, akkor minden x-re Ax-nek nullvektornak kell lennie, ez csak úgy lehet, ha A nullmátrix.
3. $$\|\lambda A\| = \sup_{x\neq 0} \frac{\|\lambda Ax\|_v}{\|x\|_v} = \sup_{x\neq 0} \frac{|\lambda| \cdot \|Ax\|_v}{\|x\|_v} = |\lambda| \cdot \sup_{x\neq 0} \frac{\|Ax\|_v}{\|x\|_v} = |\lambda| \cdot \|A\| .$$

Biz. (folytatás):

4. $$\|A+B\| = \sup_{x\neq 0} \frac{\|(A+B)x\|_v}{\|x\|_v} \le \sup_{x\neq 0} \frac{\|Ax\|_v + \|Bx\|_v}{\|x\|_v} \le$$
$$\le \sup_{x\neq 0} \frac{\|Ax\|_v}{\|x\|_v} + \sup_{x\neq 0} \frac{\|Bx\|_v}{\|x\|_v} = \|A\| + \|B\|$$

5. $B = 0 \quad \Rightarrow \quad \|B\| = 0$, valamint
$A \cdot B = A \cdot 0 = 0 \quad \Rightarrow \quad \|AB\| = 0$.
Az egyenlőtlenség mindkét oldalán 0 áll, tehát igaz az állítás.

Biz. (folytatás): Ha $B \neq 0$, akkor

$$\|A \cdot B\| = \sup_{x\neq 0} \frac{\|ABx\|_v}{\|x\|_v} = \sup_{x\neq 0, Bx\neq 0} \frac{\|ABx\|_v}{\|Bx\|_v} \cdot \frac{\|Bx\|_v}{\|x\|_v} \le$$
$$\le \sup_{Bx\neq 0} \frac{\|ABx\|_v}{\|Bx\|_v} \cdot \sup_{x\neq 0} \frac{\|Bx\|_v}{\|x\|_v} \le \sup_{y\neq 0} \frac{\|Ay\|_v}{\|y\|_v} \cdot \sup_{x\neq 0} \frac{\|Bx\|_v}{\|x\|_v} = \|A\| \cdot \|B\| .$$

Meggondolható, hogy a $Bx \neq 0$ feltétel nem változtatja meg a szuprémum értékét; közben bevezettük az $y := Bx$ jelölést. □

Tétel: Nevezetes mátrixnormák $(1, 2, \infty)$

A $\|.\|_p$ $(p = 1, 2, \infty)$ vektornormák által indukált mátrixnormák:

- $\|A\|_1 = \max\limits_{j=1}^{n} \sum\limits_{i=1}^{n} |a_{ij}|$ (oszlopnorma),
- $\|A\|_\infty = \max\limits_{i=1}^{n} \sum\limits_{j=1}^{n} |a_{ij}|$ (sornorma),
- $\|A\|_2 = \left(\max\limits_{i=1}^{n} \lambda_i(A^\top A) \right)^{1/2}$ (spektrálnorma).

Jel.: $\lambda_i(M)$: az M mátrix i-edik sajátértéke $(Mv = \lambda v,\ v \neq 0)$.

Definíció: spektrálsugár

Egy $A \in \mathbb{R}^{n \times n}$ mátrix *spektrálsugara* $\varrho(A) := \max\limits_{i=1}^{n} |\lambda_i(A)|$.

Megj.: A spektrálnormát a spektrálsugárral is meg tudjuk adni:

$$\|A\|_2 = \sqrt{\varrho(A^\top A)}.$$

Állítás:

Egy $A \in \mathbb{R}^{n \times n}$ szimmetrikus (önadjungált) mátrix spektrálnormája

$$\|A\|_2 = \varrho(A).$$

Állítás: spektrálsugár és norma

$$\varrho(A) \leq \|A\|$$

Biz.: Belátjuk, hogy $|\lambda| \leq \|A\|$.
(Legyen λ tetszőleges sajátérték és $v \neq 0$ a hozzátartozó sajátvektor.)

$$Av = \lambda v$$
$$Avv^\top = \lambda vv^\top$$
$$\|A\| \cdot \left\|vv^\top\right\| \geq \left\|Avv^\top\right\| = \left\|\lambda vv^\top\right\| = |\lambda| \cdot \left\|vv^\top\right\|$$

Leosztva $\left\|vv^\top\right\| \neq 0$-val $\|A\| \geq |\lambda|$. $\square$

10. Mátrixnormák és tulajdonságaik II.

(c) Vezesse le az 1-es vektornorma által indukált mátrixnorma képletét.

Tétel: Nevezetes mátrixnormák $(1, 2, \infty)$

A $\|.\|_p$ $(p = 1, 2, \infty)$ vektornormák által indukált mátrixnormák:

- $\|A\|_1 = \max_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}|$ (oszlopnorma),
- $\|A\|_\infty = \max_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|$ (sornorma),
- $\|A\|_2 = \left(\max_{i=1}^{n} \lambda_i(A^\top A) \right)^{1/2}$ (spektrálnorma).

Jel.: $\lambda_i(M)$: az M mátrix i-edik sajátértéke $(Mv = \lambda v,\ v \neq 0)$.

A bizonyítás „dallama":

- Az adott $f(A)$ értékre: $\|Ax\|_v \leq f(A) \cdot \|x\|_v$.
- Van olyan x vektor, hogy $\|Ax\|_v = f(A) \cdot \|x\|_v$.
- Ekkor az $f(A)$ érték, tényleg a $\|.\|_v$ vektornorma által indukált mátrixnorma, ezért jelölhetjük így: $\|A\|_v$.

Bizonyítás $\|.\|_1$ esetén: Állítás: $\|A\|_1 = \max_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}|$.

$$\|Ax\|_1 = \sum_{i=1}^{n} |(Ax)_i| = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}| \cdot |x_j| =$$

$$= \sum_{j=1}^{n} \left(|x_j| \cdot \underbrace{\sum_{i=1}^{n} |a_{ij}|} \right) \leq \underbrace{\left(\max_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}| \right)} \cdot \|x\|_1 .$$

Legyen $x = e_k$, ahol a k-adik oszlopösszeg maximális. Ekkor

$$\|Ae_k\|_1 = \underbrace{\ldots} \underbrace{\|e_k\|_1}_{1} .$$

11. LER érzékenysége.

b) Vizsgálja LER megoldásának érzékenységét szorzatfelbontások (LU, QR) alkalmazása esetén. Bizonyítsa a LER jobboldalának megváltozására vonatkozó tételt.

Példa

Hogyan befolyásolja az LU-felbontás a feladat kondicionáltságát? Mutassuk meg, hogy nem javul.

Biz.:

- $Ax = b \quad \Rightarrow \quad LUx = b \quad \Rightarrow \quad Ly = b, Ux = y,$
- $A = L \cdot U \quad \Rightarrow \quad \|A\| \le \|L\| \cdot \|U\|$
- $A^{-1} = U^{-1} \cdot L^{-1} \quad \Rightarrow \quad \|A^{-1}\| \le \|L^{-1}\| \cdot \|U^{-1}\|$
- $\mathrm{cond}\,(A) \le \mathrm{cond}\,(L) \cdot \mathrm{cond}\,(U)$ $\square$

Sőt előfordulhat, hogy $\mathrm{cond}\,(L), \mathrm{cond}\,(U) >> \mathrm{cond}\,(A)$, azaz bizonyos mátrixok esetén előfordulhat, hogy a Gauss-elimináció nagyon pontatlan eredményt ad.

Tétel: LER érzékenysége a jobb oldal pontatlanságára

Ha A invertálható és $b \neq 0$, akkor illeszkedő normákban

$$\frac{1}{\|A\| \cdot \|A^{-1}\|} \cdot \frac{\|\Delta b\|}{\|b\|} \le \frac{\|\Delta x\|}{\|x\|} \le \|A\| \cdot \left\|A^{-1}\right\| \cdot \frac{\|\Delta b\|}{\|b\|},$$

azaz

$$\frac{1}{\mathrm{cond}\,(A)} \cdot \delta b \le \delta x \le \mathrm{cond}\,(A) \cdot \delta b.$$

Biz.:

❶ $A(x + \Delta x) = (b + \Delta b)$-ből vonjuk ki az $Ax = b$ LER-t, így $A\Delta x = \Delta b$.

❷ Viszont $x = A^{-1}b$ és $\Delta x = A^{-1}\Delta b$ is teljesül.

Biz. (folytatás):

❸ Tehát a 4-féle alak:

$$b = Ax, \;\; x = A^{-1}b, \;\; \Delta b = A\Delta x, \;\; \Delta x = A^{-1}\Delta b.$$

❹ Bármely egyenlőségnél vehetjük a normát.
(A vektornormához illeszkedő mátrixnormát használunk.)

(a) $\|b\| = \|Ax\| \quad \Rightarrow \quad \|b\| \le \|A\| \cdot \|x\| \quad \Rightarrow \quad \|x\| \ge \frac{\|b\|}{\|A\|},$
(b) $\|\Delta b\| = \|A\Delta x\| \quad \Rightarrow \quad \|\Delta b\| \le \|A\| \cdot \|\Delta x\| \quad \Rightarrow \quad \|\Delta x\| \ge \frac{\|\Delta b\|}{\|A\|},$
(c) $\|x\| = \left\|A^{-1}b\right\| \quad \Rightarrow \quad \|x\| \le \left\|A^{-1}\right\| \cdot \|b\|,$
(d) $\|\Delta x\| = \left\|A^{-1}\Delta b\right\| \quad \Rightarrow \quad \|\Delta x\| \le \left\|A^{-1}\right\| \cdot \|\Delta b\|.$

❺ Az alsó becslés (b) és (c) alapján:

$$\frac{\|\Delta x\|}{\|x\|} \ge \frac{\frac{\|\Delta b\|}{\|A\|}}{\|A^{-1}\| \cdot \|b\|} = \frac{1}{\|A\| \cdot \|A^{-1}\|} \cdot \frac{\|\Delta b\|}{\|b\|}.$$

Biz. (folytatás):

6. A felső becslés (a) $\|x\| \geq \frac{\|b\|}{\|A\|}$ és (d) $\|\Delta x\| \leq \|A^{-1}\| \cdot \|\Delta b\|$ alapján:

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\|A^{-1}\| \cdot \|\Delta b\|}{\frac{\|b\|}{\|A\|}} = \|A\| \cdot \left\|A^{-1}\right\| \cdot \frac{\|\Delta b\|}{\|b\|}.$$

$\square$

12. Iterációs módszerek konvergenciája.

c) Igazolja a konvergencia szükséges és elégségséges feltételét.

Következmény: iteráció konvergenciájának elégséges feltétele

Ha $\|B\| < 1$, az $x^{(k+1)} = B \cdot x^{(k)} + c$ iteráció konvergens minden kezdőértékre.

Megj.: Attól még lehet konvergens valamely kezdőértékből indítva, ha $\|B\| \geq 1$.
(Nem szükséges feltétel.)

Lemma: spektrálsugár és az indukált normák kapcsolata

$$\varrho(B) = \inf \{ \|B\| \ : \ \|.\| \text{ indukált mátrixnorma} \},$$

azaz $\forall \, \varepsilon > 0 : \exists$ indukált $\|.\| : \|B\| < \varrho(B) + \varepsilon$.

Tétel: iteráció konvergenciájának ekvivalens feltétele

Az $x^{(k+1)} = B \cdot x^{(k)} + c$ iteráció akkor és csak akkor konvergens minden kezdőértékre, ha

$$\varrho(B) < 1.$$

Biz.:

- $\Leftarrow$: Az előző Lemma alapján trivi.
- $\Rightarrow$: Indirekt tegyük fel, hogy $\varrho(B) \geq 1$, azaz $\exists |\lambda| \geq 1$ sajátérték, és legyen $x^{(0)}$ olyan, hogy $x^{(0)} - x^*(\neq 0)$ kezdeti hiba a B λ-hoz tartozó sajátvektora legyen.

Ekkor:

$$\begin{aligned}
B(x^{(0)} - x^*) &= \lambda(x^{(0)} - x^*) \\
B^2(x^{(0)} - x^*) &= \lambda^2(x^{(0)} - x^*) \quad \Rightarrow \dots \\
B^k(x^{(0)} - x^*) &= \lambda^k(x^{(0)} - x^*) \ (k \in \mathbb{N}) \\
x^{(k)} - x^* = (Bx^{(k-1)} + c) - (Bx^* + c) &= B(x^{(k-1)} - x^*) = \\
&= B^k(x^{(0)} - x^*) = \lambda^k(x^{(0)} - x^*) \\
\left\|x^{(k)} - x^*\right\| &= |\lambda|^k \cdot \underbrace{\left\|x^{(0)} - x^*\right\|}_{\text{konst.}} \not\to 0 \quad (k \to \infty)
\end{aligned}$$

Ellentmondásra jutottunk. $\square$

13. A Jacobi-iteráció.

a) Vezesse le a Jacobi-iteráció mátrixos és koordinátás alakját. Ismertesse a csillapított változat alapötletét, határozza meg vektoros és koordinátás képleteit.

b) Írja át az iterációt a reziduumvektoros alakra, térjen ki annak szerepére. Adjon elégséges feltételt a Jacobi-iteráció konvergenciájára.

Átalakítás:

$$\begin{aligned} Ax &= b \\ (L+D+U)x &= b \\ Dx &= -(L+U)x + b \\ x &= -D^{-1}(L+U)x + D^{-1}b \end{aligned}$$

Ezek alapján az iteráció a következő.

Definíció: Jacobi-iteráció

$$x^{(k+1)} = \underbrace{-D^{-1}(L+U)}_{B_J} \cdot x^{(k)} + \underbrace{D^{-1}b}_{c_J} = B_J \cdot x^{(k)} + c_J$$

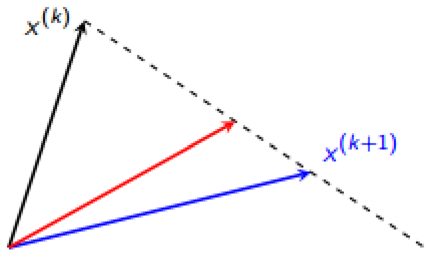
Írjuk fel koordinátánként (komponensenként)!

Állítás: a Jacobi-iteráció komponensenkénti alakja

$$x_i^{(k+1)} = \frac{-1}{a_{ii}} \left(\sum_{j=1,\, j\neq i}^{n} a_{ij} x_j^{(k)} - b_i \right) \quad (i = 1, \dots, n)$$

A *csillapítás* avagy *tompítás* alapötlete:

$$x_J^{(k+1)} \quad \text{helyett} \quad (1-\omega) \cdot x^{(k)} + \omega \cdot x_J^{(k+1)}$$



Megj.:

- alulrelaxálás ($0 < \omega < 1$), túlrelaxálás ($\omega > 1$)
- $\omega = 1$ az eredeti módszert adja

Induljunk a Jacobi-módszerből és a „helyben hagyásból":

$$\begin{array}{rcll} x & = & -D^{-1}(L+U)\cdot x + D^{-1}b & / \cdot \omega \\ x & = & x & / \cdot (1-\omega) \end{array}$$

A kettő súlyozott összege:

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

Ezek alapján az iteráció a következő.

Definíció: csillapított Jacobi-iteráció ω paraméterrel – $J(\omega)$

$$x^{(k+1)} = \underbrace{\left[(1-\omega)I - \omega D^{-1}(L+U)\right]}_{B_{J(\omega)}}\cdot x^{(k)} + \underbrace{\omega D^{-1}b}_{c_{J(\omega)}}$$

Írjuk fel koordinátánként!

Állítás: $J(\omega)$ komponensenkénti alakja

$$x_i^{(k+1)} = (1-\omega)\cdot x_i^{(k)} + \omega\cdot x_{i,J}^{(k+1)},$$

ahol $x_{i,J}^{(k+1)}$ a hagyományos Jacobi-módszer ($J = J(1)$) által adott, azaz

$$x_{i,J}^{(k+1)} = \frac{-1}{a_{i,i}}\left(\sum_{j=1,\, j\neq i}^{n} a_{i,j}x_j^{(k)} - b_i\right).$$

Írjuk fel az iteráció reziduum vektoros alakját!

$$\begin{aligned} x^{(k+1)} &= -D^{-1}(L+U)\cdot x^{(k)} + D^{-1}b = D^{-1}\left((D-A)\cdot x^{(k)} + b\right) = \\ &= x^{(k)} + D^{-1}\left(-Ax^{(k)} + b\right) = x^{(k)} + D^{-1}r^{(k)} \end{aligned}$$

Vezessük be az $s^{(k)} := D^{-1}r^{(k)}$ segédvektort, ezzel egy lépésünk alakja:

$$x^{(k+1)} = x^{(k)} + s^{(k)}.$$

Az új reziduum vektor:

$$r^{(k+1)} = b - Ax^{(k+1)} = b - A(x^{(k)} + s^{(k)}) = r^{(k)} - As^{(k)}.$$

Algoritmus: Jacobi-iteráció

$$\begin{aligned} r^{(0)} &:= b - Ax^{(0)} \\ k = 1, \ldots, \text{ leállásig} & \\ s^{(k)} &:= D^{-1} r^{(k)} \quad \Leftrightarrow \quad Ds^{(k)} = r^{(k)} \text{ LER} \\ x^{(k+1)} &:= x^{(k)} + s^{(k)} \\ r^{(k+1)} &:= r^{(k)} - As^{(k)} \end{aligned}$$

Megj.: Látjuk, hogy $x^{(k+1)} - x^{(k)} = s^{(k)}$, vagyis a tapasztalati kontrakciós együtthatók számításához lépésenként egy norma értéket és egy osztást kell elvégezni.

Tétel

Ha A szig. diag. dom. a soraira, akkor az $Ax = b$ LER-re felírt Jacobi-iteráció konvergens bármely $x^{(0)}$ esetén.

Biz.: Írjuk fel a B_J mátrix elemeit: $b_{ii} = 0$ és $i \neq j$-re $b_{ij} = -\frac{a_{ij}}{a_{ii}}$.

$$\|B_J\|_\infty = \left\| -D^{-1}(L+U) \right\|_\infty = \max_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|}$$

Ha A szig. diag. dom. a soraira, akkor

$$\forall\, i: \quad |a_{ii}| > \sum_{j=1,\, j\neq i}^{n} |a_{ij}| \quad \Leftrightarrow \quad 1 > \sum_{j=1,\, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|}.$$

Tehát minden összeg egynél kisebb, így a maximumuk is, ezzel az elégséges feltétel miatt a konvergencia teljesül.

$$\|B_J\|_\infty = \max_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|} < 1$$

14. A Gauss–Seidel-iteráció.

a) Vezesse le a Gauss–Seidel-iteráció vektoros és koordinátás alakját. Ismertesse a relaxált változat alapötletét, határozza meg vektoros és koordinátás képleteit.

Átalakítás:

$$\begin{aligned} Ax &= b \\ (L+D+U)x &= b \\ (L+D)x &= -Ux+b \\ x &= -(L+D)^{-1}Ux+(L+D)^{-1}b \end{aligned}$$

Ezek alapján az iteráció a következő.

Definíció: Gauss–Seidel-iteráció

$$x^{(k+1)} = \underbrace{-(L+D)^{-1}U}_{B_S} \cdot x^{(k)} + \underbrace{(L+D)^{-1}b}_{c_S} = B_S \cdot x^{(k)} + c_S$$

Írjuk fel koordinátánként! (Kiderül, hogy „helyben” számolható.)

Állítás: a Gauss–Seidel-iteráció komponensenkénti alakja

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

Biz.: Alakítsunk át, majd gondoljunk bele a mátrixszorzásba.

$$\begin{aligned} (L+D)x^{(k+1)} &= -Ux^{(k)}+b \\ Dx^{(k+1)} &= -Lx^{(k+1)} - Ux^{(k)} + b \\ x^{(k+1)} &= -D^{-1}(Lx^{(k+1)} + Ux^{(k)} - b) \quad \square \end{aligned}$$

Induljunk a Gauss–Seidel-iteráció következő alakjából:

$$\begin{aligned} (L+D)\cdot x &= -U\cdot x + b & / \cdot \omega \\ D\cdot x &= D\cdot x & / \cdot (1-\omega) \end{aligned}$$

A kettő súlyozott összege:

$$(D+\omega L)\cdot x = [(1-\omega)D - \omega U]\cdot x + \omega b$$

$$x = (D+\omega L)^{-1}[(1-\omega)D-\omega U]\cdot x + (D+\omega L)^{-1}\omega b$$

Ezek alapján az iteráció a következő.

Definíció: relaxált Gauss–Seidel-iteráció ω paraméterrel – $S(\omega)$

$$x^{(k+1)} = \underbrace{(D+\omega L)^{-1}[(1-\omega)D-\omega U]}_{B_{S(\omega)}} \cdot x^{(k)} + \underbrace{\omega(D+\omega L)^{-1}b}_{c_{S(\omega)}}$$

Írjuk fel koordinátánként! (Kiderül, hogy „helyben” számolható.)

Állítás: $S(\omega)$ komponensenkénti alakja

$$x_i^{(k+1)} = (1-\omega)\cdot x_i^{(k)} + \omega \cdot x_{i,S}^{(k+1)},$$

ahol $x_{i,S}^{(k+1)}$ a hagyományos Seidel-módszer ($S = S(1)$) által adott, azaz

$$x_{i,S}^{(k+1)} = \frac{-1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} + \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} - b_i \right).$$

Minden k. lépés az $i = 1, 2, \ldots, n$ sorrendben számolandó.

15. A Richardson-típusú iterációk. Kerekítési hibák az iterációkban.

a) Vezesse le a Richardson-típusú iterációk képletét. Írja fel a reziduumvektoros alakot és ismertesse annak jelentőségét. Fogalmazza meg a tanult konvergenciatételt (bizonyítás nélkül).

Tekintsük az $Ax = b$ LER-t, ahol A szimmetrikus, pozitív definit mátrix és $p \in \mathbb{R}$.

$$\begin{aligned} Ax &= b \\ p \cdot Ax &= p \cdot b \\ 0 &= -pAx + pb \\ x &= x - pAx + pb = (I - pA)x + pb \end{aligned}$$

Ezek alapján az iteráció a következő.

Definíció: Richardson-iteráció p paraméterrel – $R(p)$

$$x^{(k+1)} = \underbrace{(I - pA)}_{B_{R(p)}} \cdot x^{(k)} + \underbrace{pb}_{c_{R(p)}} = B_{R(p)} \cdot x^{(k)} + c_{R(p)}$$

Írjuk fel az iteráció reziduum vektoros alakját!

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - pAx^{(k)} + pb = x^{(k)} + p \cdot \left(-Ax^{(k)} + b\right) = \\ &= x^{(k)} + pr^{(k)} \end{aligned}$$

Vezessük be az $s^{(k)} := pr^{(k)}$ segédvektort, ezzel egy lépésünk alakja:

$$x^{(k+1)} = x^{(k)} + s^{(k)}.$$

Az új reziduum vektor:

$$r^{(k+1)} = b - Ax^{(k+1)} = b - A(x^{(k)} + s^{(k)}) = r^{(k)} - As^{(k)}.$$

Algoritmus: Richardson-iteráció

$$\begin{aligned} & r^{(0)} := b - Ax^{(0)} \\ & k = 1, \ldots, \text{leállásig} \\ & \qquad s^{(k)} := pr^{(k)} \\ & \qquad x^{(k+1)} := x^{(k)} + s^{(k)} \\ & \qquad r^{(k+1)} := r^{(k)} - As^{(k)} \end{aligned}$$

Megjegyzés: Érdemes meggondolni, hogy ha az $Ax = b$ helyett a $D = \operatorname{diag}(a_{11}, \ldots, a_{nn})$ diagonális mátrix-szal a $D^{-1}Ax = D^{-1}b$ LER-re alkalmazzuk az $R(p)$ iterációt, akkor az eredeti LER-re felírt $J(p)$ csillapított Jacobi-iterációt kapjuk.

Tétel: A Richardson-iteráció konvergenciája

Ha az $A \in \mathbb{R}^{n\times n}$ mátrix szimmetrikus, pozitív definit és sajátértékeire $m = \lambda_1 \leq \cdots \leq \lambda_n = M$ teljesül, akkor $R(p)$ (azaz az $Ax = b$ LER-re felírt $p \in \mathbb{R}$ paraméterű Richardson-iteráció) pontosan a

$$p \in \left(0, \frac{2}{M}\right),$$

paraméter értékekre konvergens minden kezdővektor esetén. Az optimális paraméter $p_0 = \frac{2}{M+m}$, a hozzá kapcsolódó kontrakciós együttható pedig:

$$\varrho(B_{R(p_0)}) := \frac{M-m}{M+m} = \left\|B_{R(p_0)}\right\|_2 = q.$$

16. A részleges LU-felbontás és az ILU algoritmus.

a) Definiálja a részleges LU-felbontást és vezesse le az ILU algoritmust. Írja át reziduumvektoros alakra is. Adjon nevezetes példákat az algoritmus speciális eseteiként.

b) Vázolja a részleges LU-felbontás előállításának algoritmusát. Adjon elégséges feltételt a felbontás létezésére és egyértelműségére.

Definíció: ILU-felbontás

- Legyen J a mátrix elemek pozícióinak egy részhalmaza, mely nem tartalmazza a főátlót, azaz $(i,i) \notin J \quad \forall\, i$-re.
  A J halmazt *pozícióhalmaz*nak nevezzük.
- Az A mátrixnak a J pozícióhalmazra illeszkedő *részleges LU-felbontásán* (*ILU-felbontásán*) olyan LU-felbontást értünk, melyre $L \in \mathcal{L}_1$ és $U \in \mathcal{U}$ (tehát a szokásos alakúak), továbbá

$$\forall\,(i,j) \in J: \quad l_{ij} = 0, \quad u_{ij} = 0 \ \text{ és}$$
$$\forall\,(i,j) \notin J: \quad a_{ij} = (LU)_{ij}.$$

Algoritmus: *ILU*-felbontás GE-val

$\widetilde{A}_1 := A$

$k = 1, \ldots, n-1:$

(1) Szétbontás: $\widetilde{A}_k = P_k - Q_k$ alakra, ahol

$$\begin{aligned}
(P_k)_{ik} &= 0 \quad (i,k) \in J\\
(P_k)_{kj} &= 0 \quad (k,j) \in J\\
(Q_k)_{ik} &= -\widetilde{a}_{ik}^{(k)} \quad (i,k) \in J\\
(Q_k)_{kj} &= -\widetilde{a}_{kj}^{(k)} \quad (k,j) \in J.
\end{aligned}$$

Ahogy látható, $\widetilde{A}_k$-nak csak k. sorában és k. oszlopában a pozícióhalmazban megadott helyeken változtatunk.

(2) Elimináció P_k-n:

$$\widetilde{A}_{k+1} = L_k P_k$$

Tétel: az *ILU*-felbontásról

Az *ILU*-felbontás algoritmusával kapott részmátrixokból készítsük el a következőket:

$$\begin{aligned}
U &:= \widetilde{A}_n,\\
L &:= L_1^{-1} \cdot \ldots \cdot L_{n-1}^{-1} \quad \text{(összepakolással)},\\
Q &:= Q_1 + Q_2 + \ldots + Q_{n-1} \quad \text{(összepakolással)}.
\end{aligned}$$

Ekkor $A = LU - Q$ és a részleges LU-felbontásra vonatkozó feltételek teljesülnek.

Biz.: A GE $n-1$. lépése után felsőháromszög alakot kapunk, tehát $U := \widetilde{A}_n$ alakja jó és minden $(i,j) \in J, i < j$-re $u_{ij} = 0$. Alkalmazzuk az $n-1$. lépés (2), majd (1) részét:

$$U := \widetilde{A}_n = L_{n-1} P_{n-1} = L_{n-1}\left(\widetilde{A}_{n-1} + Q_{n-1}\right)$$

Az $\widetilde{A}_n$-re kapott rekurziót alkalmazzuk $\widetilde{A}_{n-1}$-re:

$$\widetilde{A}_n = L_{n-1}\left(\widetilde{A}_{n-1} + Q_{n-1}\right) = L_{n-1}\left(L_{n-2}\left[\widetilde{A}_{n-2} + Q_{n-2}\right] + Q_{n-1}\right)$$

Mivel Q_{n-1}-ben az $n-2$. sorban csak nullák vannak, így az $n-2$. GE-s lépés nem változtat rajta, tehát $L_{n-2}Q_{n-1} = Q_{n-1}$. Emiatt Q_{n-1}-et bevihetjük a belső zárójelbe.

$$\widetilde{A}_n = L_{n-1}L_{n-2}\left(\widetilde{A}_{n-2} + Q_{n-2} + Q_{n-1}\right)$$

Biz. folyt.: Folytatva tovább visszafelé a rekurziót

$$U = \widetilde{A}_n = L_{n-1}L_{n-2}\left(\widetilde{A}_{n-2} + Q_{n-2} + Q_{n-1}\right) = \ldots =$$
$$= \underbrace{L_{n-1}L_{n-2}\ldots L_1}_{L^{-1}}\left(A + \underbrace{Q_1 + \ldots + Q_{n-2} + Q_{n-1}}_{Q}\right).$$

$$U = L^{-1}(A+Q) \quad \Leftrightarrow \quad A = LU - Q$$

A kapott mátrixok alakja megfelelő. Az algoritmus (1) lépése garantálja, hogy $\forall\, (i,j) \in J: \quad l_{ij} = 0, \quad u_{ij} = 0$, továbbá (2) lépése (GE) miatt $\forall\, (i,j) \notin J: \quad a_{ij} = (LU)_{ij}$. □

Tétel: szig.diag.dom. mátrix ILU-felbontása

Ha A szigorúan diagonálisan domináns a soraira vagy oszlopaira, akkor a mátrix ILU-felbontása létezik és egyértelmű.

Biz.: az ILU-felbontás (1) lépése a szig. diag. dom. tulajdonságot nem változtatja, mivel átlón kívüli elemet veszünk ki a mátrixból.

A (2) GE-s lépés a szig. diag. dom. tulajdonságot magtartja, lásd GE megmaradási tételek a Schur-komplementerre. □

Definíció: ILU-felbontás

- Legyen J a mátrix elemek pozícióinak egy részhalmaza, mely nem tartalmazza a főátlót, azaz $(i, i) \notin J \quad \forall\, i$-re.
  A J halmazt *pozícióhalmaz*nak nevezzük.
- Az A mátrixnak a J pozícióhalmazra illeszkedő *részleges LU-felbontásán* (*ILU-felbontásán*) olyan LU-felbontást értünk, melyre $L \in \mathcal{L}_1$ és $U \in \mathcal{U}$ (tehát a szokásos alakúak), továbbá

$$\forall\,(i,j) \in J: \quad l_{ij} = 0, \;\; u_{ij} = 0 \;\text{ és}$$
$$\forall\,(i,j) \notin J: \quad a_{ij} = (LU)_{ij}.$$

Átalakítás:

$$\begin{aligned}
Ax &= b, \quad A = P - Q, \quad P = LU \\
(P-Q)x &= b \\
Px &= Qx + b \\
x &= P^{-1}Qx + P^{-1}b
\end{aligned}$$

Ezek alapján az iteráció a következő.

Definíció: ILU-algoritmus

$$x^{(k+1)} = \underbrace{P^{-1}Q}_{B_{ILU}} \cdot x^{(k)} + \underbrace{P^{-1}b}_{c_{ILU}} = B_{ILU} \cdot x^{(k)} + c_{ILU}$$

Írjuk fel az iteráció reziduum vektoros alakját!

$$\begin{aligned}
A = P - Q \quad &\Leftrightarrow \quad Q = P - A \\
P \cdot x^{(k+1)} &= Q \cdot x^{(k)} + b = (P - A) \cdot x^{(k)} + b = \\
&= P \cdot x^{(k)} + \left(-Ax^{(k)} + b\right) = P \cdot x^{(k)} + r^{(k)} \\
\Rightarrow \quad x^{(k+1)} &= x^{(k)} + P^{-1}r^{(k)}
\end{aligned}$$

Vezessük be az $s^{(k)} := P^{-1}r^{(k)}$ segédvektort, ezzel egy lépésünk alakja:

$$x^{(k+1)} = x^{(k)} + s^{(k)}.$$

Az új reziduum vektor:

$$r^{(k+1)} = b - Ax^{(k+1)} = b - A(x^{(k)} + s^{(k)}) = r^{(k)} - As^{(k)}.$$

Algoritmus: *ILU*-algoritmus

$$r^{(0)} := b - Ax^{(0)}$$

$k = 1, \dots,$ leállásig

$$s^{(k)} := P^{-1} r^{(k)} \text{ helyett}$$

$$LU\, s^{(k)} = r^{(k)} \text{ (2 db háromszögű LER mo.)}$$

$$x^{(k+1)} := x^{(k)} + s^{(k)}$$

$$r^{(k+1)} := r^{(k)} - As^{(k)}$$

1. Példa:

$$P = \begin{bmatrix} 4 & 0 & 2 \\ 1 & 4 & \frac{1}{2} \\ 2 & 1 & 4 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}, \quad B_{ILU} = P^{-1}Q,$$

$$\|B_{ILU}\|_2 \approx 0.3601, \quad \|B_{ILU}\|_\infty \approx 0.3438$$

2. Példa: Jacobi-iteráció

$$P = 4\,I, \quad Q = \begin{bmatrix} 0 & -1 & -2 \\ -1 & 0 & -1 \\ -2 & -1 & 0 \end{bmatrix}, \quad B_{ILU} = P^{-1}Q = \frac{1}{4}\begin{bmatrix} 0 & -1 & -2 \\ -1 & 0 & -1 \\ -2 & -1 & 0 \end{bmatrix}$$

$$\|B_{ILU}\|_2 \approx 0.6830, \quad \|B_{ILU}\|_\infty \approx 0.75$$

3. Példa: Gauss–Seidel-iteráció

$$P = \begin{bmatrix} 4 & 0 & 0 \\ 1 & 4 & 0 \\ 2 & 1 & 4 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & -1 & -2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_{ILU} = P^{-1}Q,$$

$$\|B_{ILU}\|_2 \approx 0.6408, \quad \|B_{ILU}\|_\infty \approx 0.75$$

Látjuk, hogy az 1. példabeli *ILU*-felbontást alkalmazó *ILU*-algoritmus a leggyorsabb a három közül.

17. Nemlineáris egyenletek megoldása I.

b) Kontrakció fogalma $[a;b]$ intervallumon és a Banach-féle fixponttétel (bizonyítás nélkül). Igazolja az elégséges feltételt a kontrakcióra.

Definíció: kontrakció

A $\varphi : [a;b] \to \mathbb{R}$ leképezés *kontrakció* $[a;b]$-n, ha $\exists\, q \in [0,1)$, hogy

$$|\varphi(x) - \varphi(y)| \leq q \cdot |x - y|, \qquad \forall x, y \in [a;b].$$

Állítás

1. $\varphi\colon [a;b] \to \mathbb{R}$ függvény, $\varphi \in C^1[a;b]$ és
2. $|\varphi'(x)| < 1$ $(\forall\, x \in [a;b])$,

akkor φ kontrakció $[a;b]$-n.

Megj.:

- C^1: egyszer folyonosan differenciálható, vagyis a deriváltja folytonos.
- A kontrakciós tulajdonság függ az intervallumtól.

Biz.: A Lagrange-féle középértéktétel segítségével.

$$q := \max_{x \in [a;b]} |\varphi'(x)| < 1$$

$\forall x, y \in [a;b]$ $(x < y)$: $\exists \xi \in (x;y)$:

$$|\varphi(x) - \varphi(y)| = |\varphi'(\xi)| \cdot |x - y| \leq q \cdot |x - y|.$$

$\square$

Tétel: Banach-féle fixponttétel $[a;b]$-re

Ha a $\varphi\colon [a;b] \to [a;b]$ függvény kontrakció $[a;b]$-n q kontrakciós együtthatóval, akkor

1. $\exists!\, x^* \in [a;b] : x^* = \varphi(x^*)$, azaz létezik fixpont,
2. $\forall\, x_0 \in [a;b]$ esetén az $x_{k+1} = \varphi(x_k)$, $k \in \mathbb{N}_0$ sorozat konvergens és $\lim\limits_{k\to\infty} x_k = x^*$,
3. továbbá a következő hibabecslések teljesülnek:
   - $|x_k - x^*| \leq q^k \cdot |x_0 - x^*| \leq q^k(b-a)$,
   - $|x_k - x^*| \leq \dfrac{q^k}{1-q} \cdot |x_1 - x_0|$.

Következmény: iteráció konvergenciájának elégséges feltétele

1. Ha $\varphi\colon [a;b] \to [a;b]$,
2. $\varphi \in C^1[a;b]$ és
3. $|\varphi'(x)| < 1 \quad \forall\, x \in [a;b]$,

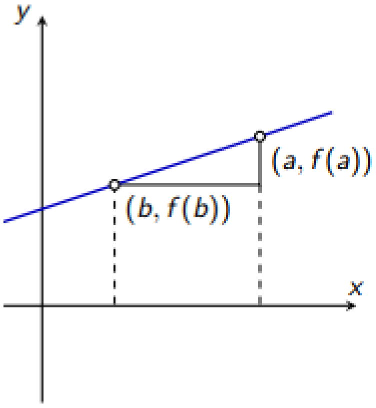
akkor az $x_{k+1} = \varphi(x_k)$ iteráció konvergens $\forall x_0 \in [a;b]$ esetén.

18. Nemlineáris egyenletek megoldása II.

a) Vázolja az intervallumfelezés algoritmusát és mutasson hozzá hibabecslést. Ismertesse a húrmódszer alapötletét, szemléltesse működését és vezesse le az algoritmusát.

b) Ismertesse a Newton-módszer alapötletét, szemléltesse működését és vezesse le a képletét. Mutassa be a többváltozós esetet is. Milyen tételt ismer a módszer monoton konvergenciájáról (bizonyítás nélkül)?

Ismétlés: Két adott ponton átmenő egyenes egyenlete.



Az egyenes meredeksége:

$$\frac{f(a)-f(b)}{a-b}.$$

Az egyenes egyenlete:

$$y-f(a)=\frac{f(a)-f(b)}{a-b}\cdot(x-a).$$

Ennek zérushelye ($y=0$):

$$x=a-\frac{f(a)\cdot(a-b)}{f(a)-f(b)}.$$

Írja le az intervallum-felezés algoritmusát és hibabecslését!

$$x_0:=a,\quad y_0:=b$$

$k=0,1\ldots$ leállásig

$$s_k:=\frac{x_k+y_k}{2}$$

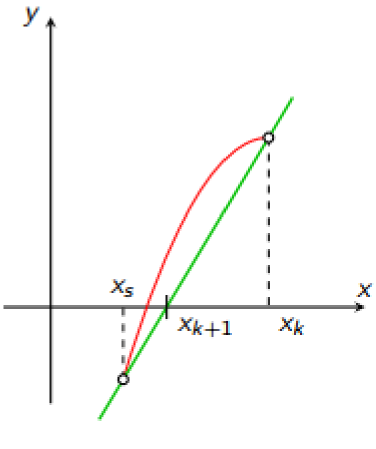
$$f(s_k)f(x_k)<0 \quad\Rightarrow\quad x_{k+1}:=x_k,\ \ y_{k+1}:=s_k$$

$$f(s_k)f(x_k)>0 \quad\Rightarrow\quad x_{k+1}:=s_k,\ \ y_{k+1}:=y_k$$

$$f(s_k)f(x_k)=0 \quad\Rightarrow\quad x^*:=\frac{x_k+y_k}{2}$$

Hibabecslés:

$$|x_k-x^*|,|y_k-x^*|<y_k-x_k\le\frac{b-a}{2^k}$$



Definíció: húrmódszer

Az $f\in C[a;b]$ függvény esetén, ha $f(a)\cdot f(b)<0$, akkor a *húrmódszer* alakja:

$$x_0:=a,\quad x_1:=b,$$

$$x_{k+1}=x_k-\frac{f(x_k)\cdot(x_k-x_s)}{f(x_k)-f(x_s)}$$

$$(k=0,1,2,\ldots),$$

ahol s a legnagyobb olyan index, amelyre $f(x_k)\cdot f(x_s)<0$.

Tétel: a húrmódszer konvergenciája

Ha $f \in C^2[a;b]$ és

1. $f(a) \cdot f(b) < 0$,
2. $M \cdot (b-a) < 1$,

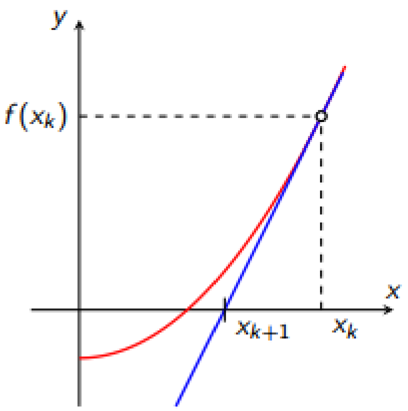
akkor a húrmódszer elsőrendben konvergál az x^* gyökhöz és

$$|x_k - x^*| \leq \frac{1}{M} \cdot (M \cdot |x_0 - x^*|)^k$$

teljesül, ahol $M = \frac{M_2}{2 \cdot m_1}$ ugyanúgy, mint korábban.

Geometriai megközelítés:

$$f, x_k \to \text{érintő} \to \text{zérushely (y=0)} \to x_{k+1}$$



Az érintő egyenlete:

$$y - f(x_k) = f'(x_k) \cdot (x - x_k)$$
$$-f(x_k) = f'(x_k) \cdot (x_{k+1} - x_k)$$
$$-\frac{f(x_k)}{f'(x_k)} = x_{k+1} - x_k$$
$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Analitikus megközelítés:

$$f \text{ gyöke} \approx x_k \text{ körüli Taylor-polinomának gyöke}$$

$$0 = f(x) = f(x_k) + f'(x_k) \cdot (x - x_k) + \dots$$

Definíció: Newton-módszer

Adott $f : \mathbb{R} \to \mathbb{R}$ differenciálható függvény és $x_0 \in \mathbb{R}$ kezdőpont esetén a *Newton-módszer* alakja:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \qquad (k = 0, 1, 2, \dots).$$

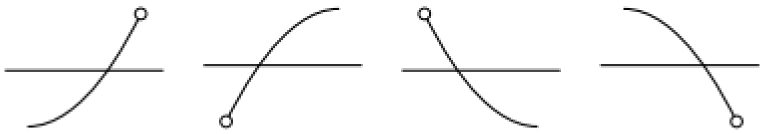
Tétel: monoton konvergencia tétele

Ha $f \in C^2[a;b]$ és

1. $\exists x^* \in [a;b] : f(x^*) = 0$, azaz van gyök,
2. f' és f'' állandó előjelű,
3. $x_0 \in [a;b] : f(x_0) \cdot f''(x_0) > 0$,

akkor az x_0 pontból indított Newton-módszer (által adott (x_k) sorozat) monoton konvergál x^*-hoz.

Megj.: 4 eset van:



Feladat

$$F : \mathbb{R}^n \to \mathbb{R}^n, \quad F(x) = 0, \qquad x = ?, \quad (x \in \mathbb{R}^n)$$

Legtöbb módszerünk általánosítható többváltozós esetre.

Egyszerű iteráció

$$F(x) = 0 \quad \Longleftrightarrow \quad x = \Phi(x)$$

Banach-féle fixponttétel szerint...

Többváltozós Newton-módszer

Közelítsük F-et az elsőfokú Taylor-polinomjával.

$$F(x) \approx F(x^{(k)}) + F'(x^{(k)}) \cdot (x - x^{(k)}),$$
$$F'(x^{(k)}) = \left(\frac{\partial f_i(x^{(k)})}{\partial x_j} \right)_{i,j=1}^{n} \in \mathbb{R}^{n \times n}$$

Ezen közelítés zérushelye lesz $x^{(k+1)}$:

1. $F'(x^{(k)}) \cdot \underbrace{(x^{(k+1)} - x^{(k)})}_{s^{(k)}} = -F(x^{(k)})$ LER megoldás ($\rightsquigarrow s^{(k)}$),
2. $x^{(k+1)} = x^{(k)} + s^{(k)}$, $s^{(k)}$ a továbblépés iránya.

Definíció: a többváltozós Newton-módszer képlete

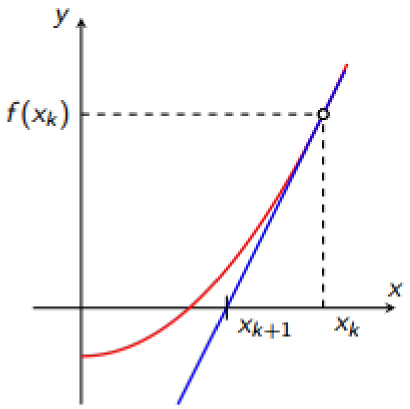
$$x^{(k+1)} = x^{(k)} - \left(F'(x^{(k)}) \right)^{-1} \cdot F(x^{(k)})$$

19. Nemlineáris egyenletek megoldása III.

a) Ismertesse a Newton-módszer alapötletét, szemléltesse működését és vezesse le a képletét. Mutassa be a többváltozós esetet is. Milyen tételt ismer a módszer lokális konvergenciájáról (bizonyítás nélkül)?

Geometriai megközelítés:

$$f, x_k \to \text{érintő} \to \text{zérushely } (y=0) \to x_{k+1}$$



Az érintő egyenlete:

$$\begin{aligned} y - f(x_k) &= f'(x_k)\cdot(x - x_k) \\ -f(x_k) &= f'(x_k)\cdot(x_{k+1} - x_k) \\ -\frac{f(x_k)}{f'(x_k)} &= x_{k+1} - x_k \\ x_{k+1} &= x_k - \frac{f(x_k)}{f'(x_k)} \end{aligned}$$

Analitikus megközelítés:

f gyöke $\approx x_k$ körüli Taylor-polinomának gyöke

$$0 = f(x) = f(x_k) + f'(x_k)\cdot(x - x_k) + \dots$$

Definíció: Newton-módszer

Adott $f\colon \mathbb{R} \to \mathbb{R}$ differenciálható függvény és $x_0 \in \mathbb{R}$ kezdőpont esetén a *Newton-módszer* alakja:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \qquad (k = 0, 1, 2, \dots).$$

Többváltozós Newton-módszer

Közelítsük F-et az elsőfokú Taylor-polinomjával.

$$F(x) \approx F(x^{(k)}) + F'(x^{(k)})\cdot(x - x^{(k)}),$$

$$F'(x^{(k)}) = \left(\frac{\partial f_i(x^{(k)})}{\partial x_j}\right)_{i,j=1}^{n} \in \mathbb{R}^{n\times n}$$

Ezen közelítés zérushelye lesz $x^{(k+1)}$:

1. $F'(x^{(k)})\cdot\underbrace{(x^{(k+1)} - x^{(k)})}_{s^{(k)}} = -F(x^{(k)})$ LER megoldás ($\rightsquigarrow s^{(k)}$),
2. $x^{(k+1)} = x^{(k)} + s^{(k)}$, $s^{(k)}$ a továbblépés iránya.

Definíció: a többváltozós Newton-módszer képlete

$$x^{(k+1)} = x^{(k)} - \left(F'(x^{(k)})\right)^{-1}\cdot F(x^{(k)})$$

Tétel: lokális konvergencia tétele

Ha $f \in C^2[a;b]$ és

1. $\exists x^* \in [a;b] : f(x^*) = 0$, azaz van gyök,
2. f' állandó előjelű,
3. $m_1 = \min_{x \in [a;b]} |f'(x)| > 0$,
4. $M_2 = \max_{x \in [a;b]} |f''(x)| < +\infty$, innen $M = \frac{M_2}{2 \cdot m_1}$.
5. $x_0 \in [a;b] : |x_0 - x^*| < r := \min\left\{ \frac{1}{M}, |x^* - a|, |x^* - b| \right\}$,

akkor az x_0 pontból indított Newton-módszer másodrendben konvergál a gyökhöz, és az

$$|x_{k+1} - x^*| \leq M \cdot |x_k - x^*|^2$$

hibabecslés érvényes.

Röviden: Ha elég közelről indulunk, akkor gyorsan odatalálunk.

Megjegyzés:

- $|x_0 - x^*| < r := \min\left\{ \frac{1}{M}, |x^* - a|, |x^* - b| \right\}$, azaz legyünk „elég közel", de azért mindenesetre legyünk $[a;b]$-n belül is.
- A monoton konvergencia feltételeinek esetén is másodrendű lesz a konvergencia, hiszen előbb-utóbb „elég közel" kerülünk a gyökhöz.

Példa:

Alkalmazzuk a következő kétváltozós függvényre a Newton-módszert!

$$F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad F : \mathbb{R}^2 \to \mathbb{R}^2,$$

ahol $f_1(x) = x_1^2 + x_2^2 - 1$, $f_2(x) = -x_1^2 - x_2$.
Geometriailag egy fordított parabola és az origó körüli egy sugarú kör metszéspontját keressük.

Megj.:

- Bizonyos pontokban a Newton-módszer nem értelmezett, mert $\det(f'(x^{(k)})) = 0$.

  $$\det(F'(x)) = \begin{vmatrix} 2x_1 & 2x_2 \\ -2x_1 & -1 \end{vmatrix} = -2x_1 + 4x_1x_2 = 2x_1(2x_2 - 1) = 0$$

  $x_1 = 0$ és $x_2 = 0.5$ esetén a módszer nem értelmezett.
- Divergens például $x_0 = \begin{bmatrix} \pm 1 & 1 \end{bmatrix}^T$-ből úgy, hogy az első koordináta sorozat konvergens (de a határérték rossz).