METODE NUMERICE – curs 13

Capitolul 1

Definiţie:

Mulţimea F, de numere în virgulă mobilă este: $F = \{x \in \mathbb{R} | x = f \cdot \beta^e\} \cup \{0\}$

$$f = \pm\left(\frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_t}{\beta^t}\right), \quad d_i \in \{0,1,\cdots,\beta-1\}, \quad i = 1,2,\cdots,t, \qquad L \le e \le U$$

- *rotunjirea uniformă*:

$$fl(x) = \begin{cases} f \cdot \beta^e, & |g| < 0,\overline{\frac{\beta}{2}} \\ f \cdot \beta^e \pm \beta^{e-t}, & |g| > 0,\overline{\frac{\beta}{2}} \\ f \cdot \beta^e \pm \beta^{e-t}, & |g| = 0,\overline{\frac{\beta}{2}}, \quad ultima\ cifra\ f - impara \\ f \cdot \beta^e, & |g| = 0,\overline{\frac{\beta}{2}}, \quad ultima\ cifra\ f - para \end{cases}$$

- Adunarea/ scăderea

Pas 1: se reprezintă intern numerele x şi y prin fl(x) şi, respectiv, fl(y):

$$fl(x) = f_x \cdot \beta^{e_x}, \quad fl(y) = f_y \cdot \beta^{e_y}$$

Pas 2: dacă $e_x \neq e_y$ → numărul cu exponent mai mic se aduce la o formă în care exponentul să fie egal cu cel al celuilalt termen → *denormalizare*

⇩

deplasarea mantisei spre dreapta, inserând zerouri după virgulă

Pas 3: se adună mantisele şi se normalizează rezultatul (dacă este necesar)

Pas 4: din rezultat se păstrează t cifre

✍ Exerciţiul 1

Care este rezultatul calculului: 8.4 - 8.35 + 1.85 realizat într-o aritmetică a virgulei mobile cu $\beta = 2$, $t = 4$, $L = -2$, $U = 4$, reprezentare cu bit ascuns, rotunjire simetrică. Se manifestă vreun fenomen de eroare? Explicaţi. Cum ar trebui efectuat calculul pentru a evita apariţia vreunui fenomen de eroare?

<u>*Rezolvare:*</u>

👉Reprezentare cu bit ascuns $\Rightarrow t = 4 + 1 = 5$

bit ascuns

$x = 8.4 = (1000.01100110011...)_2 = (0.10000\,1100110011...)_2 \times 2^4$

rotunjire simetrică $\rightarrow$ $\boxed{fl(x) = (0.10001)_2 \times 2^4}$

$y = 8.35 = (1000.01011001100...)_2 = (0.10000\,1011001100...)_2 \times 2^4$

rotunjire simetrică $\rightarrow$ $\boxed{fl(y) = (0.10001)_2 \times 2^4}$

$z = 1.85 = (1.11011001100\ldots)_2 = (0.11101\,\mathbf{1}001100\ldots)_2 \times 2^1$

rotunjire simetrică $\rightarrow$ $\boxed{fl(z) = (0.11110)_2 \times 2^1}$

$fl(x - y + z) = fl(\, fl(x - y) + z)$

$fl(x - y) = \underbrace{(0.10001)_2 \times 2^4 - (0.10001)_2 \times 2^4}_{\text{neutralizarea termenilor}} = 0$

$fl(x - y + z) = 0 + fl(z) = fl(z) = (0.11110)_2 \times 2^1$

👉 rearanjarea calculelor pentru evitarea neutralizării termenilor:

$fl(x + z - y) = fl(\, fl(x + z) - y)$

$fl(x + z) = (0.10001)_2 \times 2^4 + (0.11110)_2 \times 2^1 = (0.10001)_2 \times 2^4 + (0.00011\cancel{110})_2 \times 2^4$

denormalizare

$fl(x + z) = (0.10100)_2 \times 2^4$

$fl(x + z - y) = fl(x+z) - fl(y) = (0.10100)_2 \times 2^4 - (0.10001)_2 \times 2^4 = (0.00011)_2 \times 2^4 = (0.11000)_2 \times 2^1$

normalizare

✍ Exerciţiul 2

Care este rezultatul calculului: 5.4 + 0.15 - 5.38 realizat într-o aritmetică a virgulei mobile având $\beta = 2$, $t = 4$, reprezentare cu bit ascuns, rotunjire uniformă, $L = -2$, $U = 4$. Ce fenomen de eroare apare şi cum poate fi eliminat? Argumentaţi.

<u>Rezolvare:</u>

👉 Reprezentare cu bit ascuns $\Rightarrow$ $t = 4 + 1 = 5$

bit ascuns

$x = 5.4 = (101.01100110011...)_2 = (0.10101\,\mathbf{1}00110011...)_2 \times 2^3$

rotunjire uniformă $\rightarrow$ $\boxed{fl(x) = (0.10110)_2 \times 2^3}$

$y = 0.15 = (0.00100110011...)_2 = (0.10011\,\mathbf{0}011011001100...)_2 \times 2^{-2}$

rotunjire uniformă $\rightarrow$ $\boxed{fl(y) = (0.10011)_2 \times 2^{-2}}$

$z = 5.38 = (101.011000010...)_2 = (0.10101\,\mathbf{1}000010...)_2 \times 2^3$

rotunjire uniformă $\rightarrow$ $\boxed{fl(z) = (0.10110)_2 \times 2^3}$

$fl(x + y - z) = fl(\ fl(x + y) - z)$

$$fl(x + y) = \underbrace{(0.10110)_2 \times 2^3 + (0.10011)_2 \times 2^{-2} = (0.10110)_2 \times 2^3 + (0.00000\cancel{10011})_2 \times 2^3}_{\textit{denormalizare}} = (0.10110)_2 \times 2^3$$

omitere catastrofală

$$fl(x + y - z) = (0.10110)_2 \times 2^3 - fl(z) = \underbrace{(0.10110)_2 \times 2^3 - (0.10110)_2 \times 2^3}_{\text{neutralizarea termenilor}} = 0$$

👉 rearanjarea calculelor pentru evitarea neutralizării termenilor $\longrightarrow$ *nu este posibilă*

✍ Exerciţiul 3

Argumentaţi, pornind de la definiţie, că un algoritm instabil numeric produce, de regulă, rezultate eronate chiar dacă problema de calcul este bine condiţionată.

Rezolvare:

Notații:

D/ D - date de intrare exacte/ ușor perturbate;*

G/ G - problema de calcul/ algoritm de rezolvare a problemei de calcul*

Definiții:

Algoritm instabil $\rightarrow$ $D^* \cong D$ *astfel încât* $G(D^*) \neq G^*(D)$

Problemă bine condiționată: $D^* \cong D \Rightarrow G(D) \cong G(D^*)$

$$\Downarrow$$

$$D^* \cong D \quad \text{astfel încât} \quad G(D) \neq G^*(D)$$

Capitolul 2

☞ *tabloul general al transformărilor – triangularizarea cu pivotare parţială :*

$$M_{n-1}\cdot P_{n-1}\cdot\cdots\cdot M_2\cdot P_2\cdot M_1\cdot P_1\cdot A = U$$

- folosind $P_k = P_k^{-1}$

I_n

$$(\underbrace{M_{n-1}\cdot P_{n-1}\cdot\cdots\cdot M_1\cdot P_1\cdot P_1\cdot P_2\cdot\cdots\cdot P_{n-1}}_{(L')^{-1}})\cdot(\underbrace{P_{n-1}\cdot\cdots\cdot P_2\cdot P_1}_{P})\cdot A = U$$

matrice generală de permutare de linii

$$\boxed{P\cdot A = L'\cdot U}$$

$$L' = P_{n-1}\cdot\cdots\cdot P_2\cdot M_1^{-1}\cdot P_2\cdot M_2^{-1}\cdot\cdots\cdot P_{n-1}\cdot M_{n-1}^{-1}$$

matrice inferior triunghiulară unitate având în fiecare coloană, sub diagonala principală, subvectori Gauss cu liniile permutate

METODE NUMERICE – curs 13

□ Relaţia de recurenţă: $A = N - P$ N - matrice nesingulară

$$(N - P)\cdot \underline{x} = \underline{b} \Leftrightarrow N\cdot \underline{x} = P\cdot \underline{x} + \underline{b}$$ □ $$N\cdot \underline{x}^{[k+1]} = P\cdot \underline{x}^{[k]} + \underline{b}, \quad k = 0,1,\ldots$$

$$\underline{x}^{[k+1]} = N^{-1}\cdot P\cdot \underline{x}^{[k]} + N^{-1}\cdot \underline{b}, \quad k = 0,1,\ldots$$

<u>Notaţie</u>: $G = N^{-1}\cdot P, \quad G \in \Re^{n\times n}$

☞ *Metoda Jacobi şi metoda Gauss-Seidel* :

□ **metoda Jacobi**

$$N = D, \quad P = -(L + U)$$

□ **metoda Gauss-Seidel**

$$N = L + D, \quad P = -U$$

□ G are valori proprii în general complexe, care formează mulţimea numită <u>*spectrul matricei*</u> G

$$\rho(G) = \max_{1\le i\le n}\{|\lambda_i(G)|\}$$ - <u>*rază spectrală*</u> *a matricei* G

Teoremă:

Condiţia necesară şi suficientă ca şirul de vectori să fie convergent către soluţia sistemului de ecuaţii este ca matricea G *să aibă toate valorile proprii în modul subunitare sau, altfel spus, raza spectrală a matricei* G *să fie subunitară.*

Propoziţie:

Dacă matricea A *este diagonal dominantă pe linii, atunci metoda Jacobi este convergentă, oricare ar fi estimaţia iniţială a soluţiei sistemului de ecuaţii.*

✍ Exerciţiul 4

Calculaţi determinantul unei matrici ştiind că rezultatele factorizării sale L-U prin triangularizare cu pivotare parțială sunt:

$$L=\begin{bmatrix}1 & 0 & 0\\ 0.5 & 1 & 0\\ 0.3 & 0.54 & 1\end{bmatrix};\quad U=\begin{bmatrix}10 & -7 & 3\\ 0 & 10.1 & -4.9\\ 0 & 0 & 2.17\end{bmatrix};\quad P=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}$$

Precizaţi care este pivotul la iteraţia a doua a triangularizării şi în ce linie şi în ce coloană se află. Argumentaţi.
Rezolvați sistemul determinat de ecuații algebrice liniare având matricea de coeficienți cu rezultatele factorizarii L_U date mai sus şi coloana termenilor liberi: $\underline{b}=[1 \quad 0 \quad 1]^T$

<u>*Rezolvare:*</u>

✔ *calcul determinant*

$P \cdot A = L \cdot U \Rightarrow A = P^{-1} \cdot L \cdot U \Rightarrow \det(A) = \det(P^{-1}) \cdot \det(L) \cdot \det(U) = (-1)^{npl} \cdot \det(U)$

npl – număr de permutări de linii efectiv realizate $\rightarrow$ se determină din matricea P

$$\mathrm{P}=\mathrm{P}_2\cdot\mathrm{P}_1(\cdot\mathrm{I}_3)\Rightarrow\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\xrightarrow{\mathrm{L1}\leftrightarrow\mathrm{L2}}\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}\xrightarrow{\mathrm{L2}\leftrightarrow\mathrm{L3}}\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}=\mathrm{P}$$

☞ rezultă npl = 2 ⇒ det(A) = 10 · 10.1 · 2.17

✔ determinare pivot la iterația a 2-a:

-pivoții → pe diagonala principală a U ⇒ la iterația a 2-a: **10.1**

-la iterația a 2-a se interschimbă liniile 2 şi 3 ⇒ pivotul se afla în (linia 3, coloana 2)

✔ rezolvare sistem $A \cdot \underline{x} = \underline{b}$, $\underline{b} = [1 \quad 0 \quad 1]^T$

☞ etape:

- calcul $\underline{c} = P \cdot \underline{b} = [0 \quad 1 \quad 1]^T$
- rezolvare sistem $L \cdot \underline{y} = \underline{c}$ prin substituție înainte ⇒ $\underline{y} = [0 \quad 1 \quad 0.46]^T$
- rezolvare sistem $U \cdot \underline{x} = \underline{y}$ prin substituție înapoi ⇒ $\underline{x} = [0.078 \quad 0.202 \quad 0.212]^T$

✍ Exerciţiul 5

Fie sistemul de ecuaţii algebrice liniare cu matricea de coeficienți: $A = \begin{bmatrix} 5 & 1 \\ 7 & -2 \end{bmatrix}$

Precizaţi dacă metoda Gauss-Seidel poate fi aplicată pentru rezolvarea sistemului. Argumentaţi.

<u>*Rezolvare:*</u>

☞ verificare condiții convergență:

- suficientă $\longrightarrow$ matricea A să fie diagonal dominantă pe linii $\Rightarrow$ **nu este**

- necesară şi suficientă $\longrightarrow$ raza spectrală a matricii G = $N^{-1} \cdot P$ - subunitară

$$\mathrm{N} = \begin{bmatrix} 5 & 0 \\ 7 & -2 \end{bmatrix} \quad \Rightarrow \quad \mathrm{N}^{-1} = \begin{bmatrix} 0.2 & 0 \\ 0.7 & 0.5 \end{bmatrix}; \quad \mathrm{P} = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}; \quad \mathrm{G} = \begin{bmatrix} 0 & -0.2 \\ 0 & -0.7 \end{bmatrix}$$

$$\begin{vmatrix} \lambda & 0.2 \\ 0 & \lambda + 0.7 \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda \cdot (\lambda + 0.7) = 0 \quad \Rightarrow \quad \rho(\mathrm{G}) = 0.7 < 1$$

⇓

metoda Gauss-Seidel poate fi aplicată pentru rezolvarea sistemului

✍ Exerciţiul 6

Realizaţi funcția MATLAB ce returnează coeficienții calculați în sensul celor mai mici pătrate a ai dependenţei:

$$y(u,v,z) = a \cdot u \cdot v + b \cdot v^2 \cdot z + c \cdot (u^2 + v \cdot z)$$

pe baza valorilor cunoscute $\{u_i, v_i, z_i, y(u_i, v_i, z_i)\}_{i=\overline{1,n}}, n > 3$ *stocate într-o matrice cu n linii şi 4 coloane. Se vor utiliza funcțiile elementare ale mediului MATLAB, precum şi funcția " \ ".*

<u>*Rezolvare:*</u>

☞ formare sistem supradeterminat:

- notație: $y_i = y(u_i, v_i, z_i), \quad i = \overline{1,n}$ $\longrightarrow$ $y_i = [u_i \cdot v_i \quad v_i^2 \cdot z_i \quad u_i^2 + v_i \cdot z_i] \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad i = \overline{1,n}$

$$\Downarrow$$

$$\underbrace{\begin{bmatrix} u_1 \cdot v_1 & v_1^2 \cdot z_1 & u_1^2 + v_1 \cdot z_1 \\ u_2 \cdot v_2 & v_2^2 \cdot z_2 & u_2^2 + v_2 \cdot z_2 \\ \vdots & \vdots & \vdots \\ u_n \cdot v_n & v_n^2 \cdot z_n & u_n^2 + v_n \cdot z_n \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\underline{x}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\underline{b}}$$

☞parametrii de intrare – ieşire funcție MATLAB:

-intrare: matrice de dimensiune n x 4 (valorile cunoscute ale u, v, z, y);

-ieşire: vector de dimensiune 3 x 1 (vector <u>x</u>).

☞structura funcției:

- verificare n > 3 (n – număr de linii variabila de intrare)
- formare matrice de coeficienți sistem supradeterminat, A;
- formare vector termeni liberi sistem supradeterminat, <u>b</u>;
- rezolvare sistem utilizând “\” $\rightarrow$ parametru ieşire funcție

Exerciţiul 7

Se consideră o matrice $A \in \Re^{mxn}, m > n$, *cu structura:*

$$A = \begin{bmatrix} R & \vdots & B \\ \cdots & \vdots & \cdots \\ \mathbf{0}_{(m-k)xk} & \vdots & C \end{bmatrix}$$

unde $R \in \Re^{kxk}$ *superior triunghiulară,* $B \in \Re^{kx(n-k)}$, $C \in \Re^{(m-k)x(n-k)}$. *Ce elemente ale matricei A se modifică în urma unei transformări de tipul* $U \cdot A$, *unde* $U = I_m - \frac{\underline{u} \cdot \underline{u}^T}{\beta}$, $\underline{u}^T = [0 \ldots 0\ u_{k+1} \ldots u_m]$ *? Argumentaţi.*

<u>*Rezolvare:*</u>

$$U \cdot A = U \cdot [\underline{c}_1 \quad \ldots \quad \underline{c}_i \quad \ldots \quad \underline{c}_n], \quad \underline{c}_i \in \Re^{mx1}, i = \overline{1, n}$$

$$U \cdot \underline{c}_i = \left(I_m - \frac{\underline{u} \cdot \underline{u}^T}{\beta} \right) \cdot \underline{c}_i = \underline{c}_i - \frac{\underline{u} \cdot \underline{u}^T}{\beta} \cdot \underline{c}_i = \underline{c}_i - \tau \cdot \underline{u} = \begin{bmatrix} c_{1,i} \\ \vdots \\ c_{k,i} \\ c_{k+1,i} - u_{k+1} \cdot \tau \\ \vdots \\ c_{m,i} - u_m \cdot \tau \end{bmatrix}$$

$$\tau = \frac{\underline{u}^T}{\beta} \cdot \underline{c}_i = \sum_{j=k+1}^{m} \frac{u_j \cdot c_{j,i}}{\beta}$$

$\boxed{\text{se modifică elementele din liniile de la } k + 1 \div m}$

- pentru i = 1, ..., k $\Rightarrow$ $c_{j,i}$ = 0, j = k + 1, ..., m $\Rightarrow$ τ = 0 $\longrightarrow$ primele k coloane nu se modifică