Calcul Numeric

Cursul 2

2020

Anca Ignat

Calcul matricial

Fie matricea $\boldsymbol{A} \in \mathbb{R}^{m\times n}$:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad , \quad \boldsymbol{A} = \left(a_{ij}\right)_{i=1\ldots m,\, j=1\ldots n}$$

Se defineşte ***matricea transpusă***:

$$\boldsymbol{A}^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}, \boldsymbol{A}^T = \left(a_{ji}\right)_{i=1\ldots m,\, j=1\ldots n} \in \mathbb{R}^{n\times m}$$

Pentru matricea:

$$A \in \mathbb{C}^{m\times n},\ A = \left(a_{ij}\right)_{\substack{i=1\ldots m\\ j=1\ldots n}}$$

se defineşte ***matricea adjunctă*** A^H:

$$A^H = \overline{A^T} = \left(\overline{a_{ji}}\right)_{\substack{j=1\ldots n\\ i=1\ldots m}}$$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad A^H = \begin{pmatrix} \overline{a_{11}} & \cdots & \overline{a_{m1}} \\ \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \cdots & \overline{a_{mn}} \end{pmatrix}$$

Pentru $A \in \mathbb{R}^{m\times n}$ matricea adjunctă coincide cu transpusa,

$A^H = A^T$.

Fie vectorul $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$, acesta este considerat vector coloană, $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}\times \mathbf{1}}$:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{x}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$$

Dacă facem înmulţirea matricială $\boldsymbol{Ae_j}$ obţinem coloana $\boldsymbol{j}$ a matricei $\boldsymbol{A}$:

$$\boldsymbol{Ae_j} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1_{\text{poziţia } j} \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

$\boldsymbol{Ae_j}$ este coloana $\boldsymbol{j}$ a matricei $\boldsymbol{A}$, $\boldsymbol{j=1,...,n}$;

$\boldsymbol{e_i^T A}$ este linia $\boldsymbol{i}$ a matricei $\boldsymbol{A}$, $\boldsymbol{i=1,...,m}$.

Fie vectorii ***x, y***, cu ajutorul lor definim produsele scalare în $\mathbb{C}^n$ şi $\mathbb{R}^n$:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{C}^n \ , \ \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{C}^n$$

$$(x\,,\,y) = \sum_{i=1}^{n} x_i\,\overline{y_i} = y^H x = \begin{pmatrix} \overline{y_1} & \overline{y_2} & \cdots & \overline{y_n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \; , \; y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$$

$$(x \, , \, y) = \sum_{i=1}^{n} x_i y_i = y^T x = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Proprietăţile matricei $\boldsymbol{A^H}$ | Proprietăţi ale matricei $\boldsymbol{A^T}$

	A^H	A^T
1.	$(A + B)^H = A^H + B^H$	$(A + B)^T = A^T + B^T$
2.	$(A^H)^H = A$	$(A^T)^T = A$
3.	$(AB)^H = B^H A^H$	$(AB)^T = B^T A^T$
4.	$(A^{-1})^H = (A^H)^{-1}$	$(A^{-1})^T = (A^T)^{-1}$

Propoziţie

Fie $\boldsymbol{A} \in \mathbb{C}^{\boldsymbol{m} \times \boldsymbol{n}}$, $\boldsymbol{x} \in \mathbb{C}^{\boldsymbol{n}}$, $\boldsymbol{y} \in \mathbb{C}^{\boldsymbol{m}}$ atunci:

$$\left(\boldsymbol{Ax}, \boldsymbol{y}\right)_{\mathbb{C}^m} = \left(\boldsymbol{x}, \boldsymbol{A}^{\boldsymbol{H}} \boldsymbol{y}\right)_{\mathbb{C}^n} .$$

Pentru cazul real avem:

$$\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}, \ \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, \boldsymbol{y} \in \mathbb{R}^{\boldsymbol{m}} \Rightarrow \left(\boldsymbol{Ax}, \boldsymbol{y}\right)_{\mathbb{R}^m} = \left(\boldsymbol{x}, \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{y}\right)_{\mathbb{R}^n}$$

Demonstraţie

$$\boldsymbol{(Ax, y) = y^H (Ax) = y^H A\, x = y^H (A^H)^H x =}$$

$$\boldsymbol{= (A^H y)^H x = (x, A^H y).}$$

Tipuri de matrice

Definiţii

O matrice $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ se numeşte ***simetrică*** dacă $\boldsymbol{A} = \boldsymbol{A}^T$.

O matrice $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ se numeşte ***autoadjunctă*** dacă $\boldsymbol{A} = \boldsymbol{A}^H$.

O matrice $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ se numeşte ***unitară*** dacă $\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{A}\,\boldsymbol{A}^H = \boldsymbol{I}_n$.

O matrice $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ se numeşte ***ortogonală*** dacă

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\,\boldsymbol{A}^T = \boldsymbol{I}_n.$$

O matrice $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\boldsymbol{A=(a_{ij})}$ se numește matrice ***triunghiulară inferior*** (sau ***inferior triunghiulară***) dacă

$$\boldsymbol{a_{ij} = 0} \text{ pentru } \boldsymbol{j > i}$$

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 & 0 \\ \vdots & & & & & \\ a_{(n-1)1} & a_{(n-1)2} & a_{(n-1)3} & \cdots & a_{(n-1)(n-1)} & 0 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{n(n-1)} & a_{nn} \end{pmatrix}$$

O matrice $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\boldsymbol{A=(a_{ij})}$ se numește matrice ***triunghiulară superior*** (sau ***superior triunghiulară***) dacă

$$a_{ij} = 0 \text{ pentru } j < i$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2(n-1)} & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3(n-1)} & a_{3n} \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

Notăm cu $\boldsymbol{I_n}$ matricea unitate:

$$\boldsymbol{I_n} \in \mathbb{R}^{\boldsymbol{n \times n}}, \boldsymbol{I_n} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

Matrice diagonală $\boldsymbol{D}$**=diag**$[\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_n}]$

$$D \in \mathbb{R}^{n\times n},\ D = \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & d_2 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & d_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & d_n \end{pmatrix}$$

Norme

Definiție

Fie $\boldsymbol{X}$ un spațiu vectorial real. Se numește ***normă*** aplicația:

$$\|\ .\ \|: X \to \mathbb{R}_{+}$$

care îndeplinește condițiile:

$$\textbf{(1)}\ \ \|x\| \geq 0; \quad \|x\| = 0 \iff x = 0;$$

$$\textbf{(2)}\ \ \|x+y\| \leq \|x\| + \|y\|, \forall\ x, y \in X;$$

$$\textbf{(3)}\ \ \|\lambda x\| = |\lambda| \|x\|, \forall\ x \in X, \forall \lambda \in \mathbb{R}.$$

Vom numi ***norme vectoriale*** normele definite pe spațiile $X = \mathbb{C}^n$ sau $\mathbb{R}^n$.

Exemple

Fie spațiile vectoriale $\mathbb{C}^n$ sau $\mathbb{R}^n$. Pe aceste spații următoarele aplicații sunt norme vectoriale:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|;$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2};$$

$$\|x\|_\infty = \max\{|x_i|, i = 1..n\}.$$

Dacă $\|\cdot\|_v$ este o normă vectorială şi $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ este o matrice nesingulară atunci aplicația:

$$\|\cdot\|_P : \mathbb{R}^n \to \mathbb{R}, \qquad \|\boldsymbol{x}\|_P = \|\boldsymbol{P}\boldsymbol{x}\|_v$$

este de asemenea o normă vectorială.

Definiție

Se numește ***produs scalar*** în spațiul vectorial $\boldsymbol{X}$ aplicația:

$$(\cdot\,,\cdot): X \times X \to K$$

care satisface condițiile :

$$\textbf{(a)}\ (x\,,x) \geq 0\,,\forall\ x \in X, \quad (x,x) = 0 \ \Leftrightarrow \ x = 0;$$

$$\textbf{(b)}\ (x\,,y) = \overline{(y\,,x)}\,,\forall\ x\,,y \in X,$$

$$\textbf{(c)}\ (\lambda x\,,y) = \lambda(x\,,y),\forall\ x\,,y \in X\,,\forall \lambda \in K,$$

$$\textbf{(d)}\ (x+y\,,z) = (x\,,z) + (y\,,z),\forall\ x\,,y\,,z \in X\,.$$

Inegalitatea lui Cauchy-Buniakovski-Schwarz:

$$\left|(\boldsymbol{x}, \boldsymbol{y})\right| \leq \sqrt{(\boldsymbol{x}, \boldsymbol{x})}\sqrt{(\boldsymbol{y}, \boldsymbol{y})} \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$$

Într-un spațiu vectorial dotat cu produs scalar se poate induce o normă numită euclidiană:

$$\left\|\boldsymbol{x}\right\|_{\mathbf{2}} = |\,\boldsymbol{x}\,| := \sqrt{(\boldsymbol{x}, \boldsymbol{x})}.$$

Reamintim definiția produselor scalare pe $\mathbb{C}^n$ şi pe $\mathbb{R}^n$ introduse anterior:

$$(x\,,y)_{\mathbb{C}^n}=\sum_{i=1}^{n}x_i\,\overline{y_i}\quad,\quad(x\,,y)_{\mathbb{R}^n}=\sum_{i=1}^{n}x_i\,y_i$$

Obținem norma euclidiană (valabilă în spaţiile $\mathbb{C}^n$ şi $\mathbb{R}^n$):

$$\left\|x\right\|_2=|\,x\,|=\sqrt{\sum_{i=1}^{n}\left|x_i\right|^2}\,.$$

Norme matriceale

Definiție

Aplicația $\|\cdot\|: \mathbb{R}^{n\times n} \to \mathbb{R}$ se numește ***normă matriceală*** dacă:

$$\textbf{(1)}\ \|A\| \geq 0\ \forall\, A \in \mathbb{R}^{n\times n}\ ;\ \|A\| = 0 \Leftrightarrow A = 0.$$

$$\textbf{(2)}\ \|\alpha A\| = |\alpha|\,\|A\|\ , \forall \alpha \in \mathbb{R}\, , \forall\, A \in \mathbb{R}^{n\times n}.$$

$$\textbf{(3)}\ \|A + B\| \leq \|A\| + \|B\|, \forall\, A, B \in \mathbb{R}^{n\times n}.$$

$$\textbf{(4)}\ \|A * B\| \leq \|A\| \cdot \|B\|\, , \forall\, A, B \in \mathbb{R}^{n\times n}.$$

Exemple

Norma Frobenius definită de relația $\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}$ este o normă matriceală.

Aplicația $\|A\|_{\max} = \max\{|a_{ij}|; i = 1,\ldots,n,\ j = 1,\ldots,n\}$ ***NU*** este o normă matriceală.

Pentru ***n = 2*** fie:

$$\boldsymbol{A}=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix},\ \boldsymbol{B}=\boldsymbol{A}^{T}=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}$$

$$\boldsymbol{A}*\boldsymbol{B}=\boldsymbol{I}_2\ ,\ \|\boldsymbol{A}\|_{\max}=\|\boldsymbol{B}\|_{\max}=\frac{1}{\sqrt{2}}$$

$$\|\boldsymbol{A}*\boldsymbol{B}\|_{\max}=1>\ \|\boldsymbol{A}\|_{\max}\bullet\|\boldsymbol{B}\|_{\max}=\frac{1}{2}.$$

Norme matriceale naturale

- $\|\cdot\|_v : \mathbb{R}^n \to \mathbb{R}_+$ o normă vectorială $\to \|\cdot\|_i : \mathbb{R}^{n\times n} \to \mathbb{R}_+$ ***normă matriceală naturală*** sau ***indusă***.

$$\|A\|_{\mathrm{i}} = \max\{\frac{\|Ax\|_{\mathrm{v}}}{\|x\|_{\mathrm{v}}} ; x \in \mathbb{R}^n , x \neq 0\}$$

Definiţii echivalente :

$$\|A\|_{\mathrm{i}} = \max\{\|Ax\|_{\mathrm{v}} \ ; x \in \mathbb{R}^n , \|x\|_{\mathrm{v}} \leq 1\}$$

$$= \max\{\|Ax\|_{\mathrm{v}} \ ; x \in \mathbb{R}^n , \|x\|_{\mathrm{v}} = 1\}$$

$\|A\|_{\mathrm{i}}$ se numeşte ***normă matricealā naturală*** sau ***normă indusă*** de norma vectorială $\|\cdot\|_v$

Avem următoarea relaţie:

$$\|Ax\|_{\mathrm{v}} \le \|A\|_{\mathrm{i}}\,\|x\|_{\mathrm{v}}\,, \forall\, A \in \mathbb{R}^{n\times n}\,, \forall x \in \mathbb{R}^{n}.$$

Norma Frobenius $\|\cdot\|_F$ nu este o normă naturală.

$$\|I_n\|_{\mathrm{i}} = \max\{\frac{\|I_n x\|_{\mathrm{v}}}{\|x\|_{\mathrm{v}}}\,;\, x \ne 0\} = 1,\ \ \forall\ \|\cdot\|_i,$$

$$\|I_n\|_F = (1+1+\cdots+1)^{1/2} = \sqrt{n} \ne 1 \ \text{pentru}\ \ n \ge 2.$$

Pentru $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ norma matriceală indusă este:

$$\|A\|_1 = \max\{\sum_{i=1}^{n} |a_{ij}|;\ j = 1, 2, \ldots, n\}$$

Pentru $\|x\|_\infty = \max\{|x_i|;\ i = 1, \ldots, n\}$ norma matricială indusă este:

$$\|A\|_\infty = \max\{\sum_{j=1}^{n} |a_{ij}|;\ i = 1, 2, \ldots, n\}.$$

- $\|\cdot\|_{\mathbf{v}}$ şi $\|\cdot\|_{\mathbf{v,P}}$ - norme vectoriale → $\|\cdot\|_{\mathbf{i}}$ şi respectiv $\|\cdot\|_{\mathbf{i,P}}$

normele matriciale induse

$$\|\boldsymbol{x}\|_{\mathbf{v,P}} = \|\boldsymbol{Px}\|_{\mathbf{v}} \quad \rightarrow \quad \|\boldsymbol{A}\|_{\mathbf{i,P}} = \|\boldsymbol{PAP}^{-1}\|_{\mathbf{i}}$$

Valori şi vectori proprii

Definiţii

Fie $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n \times n}}$. Se numeşte ***valoare proprie*** (***autovaloare***) a matricei $\boldsymbol{A}$ un număr complex $\boldsymbol{\lambda} \in \mathbb{C}$ pentru care există un vector nenul $\boldsymbol{x} \in \mathbb{C}^{\boldsymbol{n}}, \boldsymbol{x} \neq \mathbf{0}$ a.î.:

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

Vectorul x se numeşte ***vector propriu*** (***autovector***) asociat val. proprii $\boldsymbol{\lambda}$.

$$\boldsymbol{Ax = \lambda x} \;\Leftrightarrow\; \boldsymbol{(\lambda I_n - A)x = 0, x \neq 0} \;\Leftrightarrow\; \mathbf{det}\boldsymbol{(\lambda I_n - A) = 0}$$

→ Matricea $\lambda I_n - A$ este singulară.

Polinomul:

$$p_A(\lambda) = \det(\lambda I_n - A) = \lambda^n - a_1\lambda^{n-1} - a_2\lambda^{n-2} - \ldots - a_{n-1}\lambda - a_n$$

se numeşte ***polinom caracteristic*** asociat matricei ***A***.

→ **grad** $p_A = n$ → are n rădăcini care sunt valorile proprii ale matricei ***A***.

Se numeşte ***rază spectrală*** a matricei ***A***:

$$\rho(A) = \max\{|\lambda_i|, i = 1,\ldots,n, \lambda_i - \text{valorile proprii ale matricei A}\}$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$ norma indusă este

$$\|A\|_2 = |A| = \sqrt{\rho(A^T A)}$$ se numeşte ***norma spectrală***.

Propoziţia 1

Fie $\|\cdot\|$ o normă matriceală naturală. Atunci:

$$\rho(A) \le \|A\|, \forall\, A \in \mathbb{R}^{n\times n}.$$

Fie $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n\times n}}$ **,$\{\boldsymbol{A^k}\}$ un şir de matrici.**

$$\boldsymbol{A^k \to 0_{n\times n}\,, k \to \infty} \;\Leftrightarrow\; \left\|\boldsymbol{A^k}\right\| \boldsymbol{\to 0\,, k \to \infty}.$$

Propoziţia 2

Fie $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n\times n}}$. Atunci:

$$\boldsymbol{A^k \to 0\,, k \to \infty} \;\Leftrightarrow\; \boldsymbol{\rho(A)} < 1.$$

Dacă există o normă matriceală naturală pentru care **||*A*|| < *1***

atunci:

$$\boldsymbol{A^k \to 0 \text{ pentru } k \to \infty.}$$

$$(\boldsymbol{n = 1 \to a} \in \mathbb{R}\,, \boldsymbol{a^k \to 0 \text{ pentru } k \to \infty} \;\Leftrightarrow\; \left|\boldsymbol{a}\right| \boldsymbol{< 1.})$$

Propoziţia 3

Fie $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n \times n}}$. Seria $\sum_{\boldsymbol{k=0}}^{\infty} \boldsymbol{A}^{\boldsymbol{k}}$ converge dacă şi numai dacă raza spectrală a matricei $\boldsymbol{A}$ este subunitară:

$$\sum_{\boldsymbol{k}=0}^{\boldsymbol{n}} \boldsymbol{A}^{k} = \boldsymbol{S} \quad \Leftrightarrow \quad \rho(\boldsymbol{A}) < \mathbf{1}.$$

Dacă există o normă a matricei A astfel încât $\|\boldsymbol{A}\| < \boldsymbol{1}$ atunci seria converge. În cazul convergenței avem :

$$\sum_{\boldsymbol{k=0}}^{\infty} \boldsymbol{A}^{\boldsymbol{k}} = \boldsymbol{S} = (\boldsymbol{I} - \boldsymbol{A})^{-1}.$$

Propoziţia 4

Fie $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ pentru care există o normă matriceală naturală astfel ca $\|\boldsymbol{A}\| < \mathbf{1}$. Atunci există matricele $(\boldsymbol{I}_n \pm \boldsymbol{A})^{-1}$ şi avem evaluările:

$$\frac{\mathbf{1}}{\mathbf{1} + \|\boldsymbol{A}\|} \leq \left\|(\boldsymbol{I} \pm \boldsymbol{A})^{-1}\right\| \leq \frac{\mathbf{1}}{\mathbf{1} - \|\boldsymbol{A}\|}.$$

Numere în format binar

În 1985 IEEE a publicat un raport numit Binary Floating Point Arithmetic Standard 754-1985 și o actualizare în 2008 IEEE 754-2008 care furnizează standarde pentru numere în virgulă mobilă binare și decimale, formate de interschimbare a tipului de date, algoritmi de rotunjire aritmetică, tratarea excepțiilor. Aceste standarde sunt respectate de toți fabricanții de calculatoare care folosesc arhitectura în virgulă mobilă.

O reprezentare binară pe 64 de biți a unui număr real se face în felul următor: primul bit este bitul de semn, următorii 11 biți reprezintă exponentul ***c*** iar următorii 52 de biți conțin informații despre partea fracționară, ***f***, numită și mantisă:

$$(-1)^s 2^{c-1023}(1+f)\ .$$

0 10000000011 1011100100010000000000000000000000000000000000000000 = 27.56640625

[27.566406249999998223643160599749535322189331054687 5, 27.5664062500000017763568394002504646778106689453125).

Cel mai mic număr pozitiv care poate fi reprezentat este cu $s=0, c=1, f=0$ adică

$$z = 2^{-1022}(1+0) \approx 0.22251 \times 10^{-307}$$

iar cel mai mare este pentru $s=0, c=2046, f=1-2^{-52}$

$$Z = 2^{1023}(2-2^{-52}) \approx 0.17977 \times 10^{309}.$$

Numerele care apar în calcule și sunt mai mici decât z sunt setate în general la 0 (*underflow*) iar cele mai mari decât Z duc, de obicei, la oprirea calculelor (*overflow*).

Se observă că numărul 0 are două reprezentări: $s=0, c=1, f=0$ și $s=1, c=1, f=0$.

Reprezentarea zecimală

$\pm 0.d_1 d_2 \ldots d_k \times 10^n$ $1 \le d_1 \le 9$, $0 \le d_i \le 9$, $i=2,...,k$ - reprezentarea zecimală folosind k cifre. Orice număr real y:

$$y = 0.d_1 d_2 \ldots d_k d_{k+1} d_{k+2} \ldots \times 10^n$$

poate fi reprezentat folosind k cifre printr-o simplă trunchiere

$$fl(y) = 0.d_1 d_2 \ldots d_k \times 10^n \quad .$$

O altă metodă de a obține o reprezentare cu ***k*** cifre este prin rotunjire:

$$\boldsymbol{fl(y) = 0.\delta_1\delta_2\ldots\delta_k \times 10^n}$$

Dacă $\boldsymbol{d_{k+1} \geq 5}$ se adaugă ***1*** la $\boldsymbol{d_k}$ pentru a obține ***fl(y)*** (*round up*),

altfel se face trunchierea la ***k*** cifre (*round down*).

Un număr $\boldsymbol{r^*}$ aproximează numărul $\boldsymbol{r}$ cu ***t*** cifre exacte dacă t este cel mai mare număr întreg nenegativ pentru care:

$$\boldsymbol{\frac{|r - r^*|}{|r|} \leq 5 \times 10^{-t}} \ .$$

În cazul trunchierii avem

$$\left|\frac{y - fl(y)}{y}\right| \leq 10^{-k+1}$$

iar când se face rotunjirea:

$$\left|\frac{y - fl(y)}{y}\right| \leq 0.5 \times 10^{-k+1}.$$

Operațiile elementare

$$x +_c y = fl(fl(x) + fl(y))$$
$$x -_c y = fl(fl(x) - fl(y))$$
$$x \times_c y = fl(fl(x) \times fl(y))$$
$$x \div_c y = fl(fl(x) \div fl(y))$$

Surse de erori în calculele numerice

1. Erori în datele de intrare:
   - măsurători afectate de erori sistematice sau perturbaţii temporare,
   - erori de rotunjire: 1/3 , π , 1/7,…
2. Erori de rotunjire în timpul calculelor:
   - datorate capacităţii limitate de memorare a datelor, operaţiile nu sunt efectuate exact.

3. Erori de discretizare:

- limita unui şir , suma unei serii , funcţii neliniare aproximate de funcţii liniare, aproximarea derivatei unei funcţii

4. Simplificări în modelul matematic

- idealizări , ignorarea unor parametri.

5. Erori umane şi erori ale bibliotecilor folosite.

Eroare absolută , eroare relativă

a – valoarea exactă,

ã – valoarea aproximativă.

Eroare absolută : ***a- ã sau |a - ã | sau*** $\|a - \tilde{a}\|$

$$a = \tilde{a} \pm \Delta_a \,,\ |a - \tilde{a}| \le \Delta_a$$

Eroare relativă: $a \neq 0$ $\quad \dfrac{a - \tilde{a}}{a}$ **sau** $\dfrac{|a - \tilde{a}|}{|a|}$ **sau** $\dfrac{\|a - \tilde{a}\|}{\|a\|}$

$\dfrac{|a - \tilde{a}|}{|a|} \le \delta_a$ (δ_a se exprimă, de regulă, în %).

În aproximările 1kg ±5g, 50g±5g erorile absolute sunt egale dar pentru prima cantitate eroarea relativă este 0,5% iar pentru a doua eroarea relativă este 10%.

$$a_1 = \tilde{a}_1 \pm \Delta_{a_1}\ , a_2 = \tilde{a}_2 \pm \Delta_{a_2}\ ,$$

$$a_1 \pm a_2 = (\tilde{a}_1 \pm \tilde{a}_2) \pm \left(\Delta_{a_1} \pm \Delta_{a_2}\right)$$

$$\Delta_{a_1 + a_2} \leq \Delta_{a_1} + \Delta_{a_2}.$$

a₁ cu eroare relativă δ_{a_1} şi ***a₂*** cu eroare relativă δ_{a_2} :

$a = a_1 * a_2$ sau $\dfrac{a_1}{a_2}$ rezultă $\delta_a = \delta_{a_1} + \delta_{a_2}$.