Calcul Numeric

Cursul 3

2017

Anca Ignat

Surse de erori în calculule 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*** $\|a-\tilde{a}\|$

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

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

$$\frac{|a-\tilde{a}|}{|a|} \le \delta_a$$ (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%.

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

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

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

a1 cu eroare relativă $\boldsymbol{\delta_{a_1}}$ şi ***a2*** cu eroare relativă $\boldsymbol{\delta_{a_2}}$:

$$\boldsymbol{a = a_1 * a_2} \text{ sau } \boldsymbol{\frac{a_1}{a_2}} \text{ rezultă } \quad \boldsymbol{\delta_a = \delta_{a_1} + \delta_{a_2}}.$$

Condiţionare ←→ stabilitate

Condiţionarea unei probleme caracterizează sensibilitatea soluţiei în raport cu perturbarea datelor de intrare, în ipoteza unor calcule exacte (independent de algoritmul folosit pentru rezolvarea problemei).

Fie $\boldsymbol{x}$ datele exacte de intrare, $\tilde{\boldsymbol{x}}$ o aproximaţie cunoscută a acestora, $\boldsymbol{P(x)}$ soluţia exactă a problemei şi $\boldsymbol{P(\tilde{x})}$ soluţia problemei cu $\tilde{\boldsymbol{x}}$ ca date de intrare. Se presupune că s-au făcut calcule exacte la obţinerea soluţiilor $\boldsymbol{P(x)}$ şi $\boldsymbol{P(\tilde{x})}$.

O problemă se consideră a fi ***prost condiţionată*** dacă ***P(x)*** şi $P(\tilde{x})$ diferă mult chiar dacă eroarea relativă $\frac{\| x - \tilde{x} \|}{\| x \|}$ este mică.

Condiţionarea numerică a unei probleme este exprimată prin amplificarea erorii relative:

$$k(x) = \frac{\dfrac{\| P(x) - P(\tilde{x}) \|}{\| P(x) \|}}{\dfrac{\| x - \tilde{x} \|}{\| x \|}} \quad \textbf{pentru } x \neq 0 \ \textbf{ si } P(x) \neq 0$$

O valoare mică pentru ***k(x)*** caracterizează o problemă bine-condiţionată.

Condiţionarea este o proprietate locală (se evaluează pentru diverse date de intrare $\boldsymbol{x}$). O problemă este bine-condiţionată dacă este bine-condiţionată în orice punct.

Pentru rezolvarea unei probleme $\boldsymbol{P}$, calculatorul execută un algoritm $\tilde{\boldsymbol{P}}$. Deoarece se folosesc numere în virgulă mobilă, calculele sunt afectate de erori:

$$\boldsymbol{P(x) \neq \tilde{P}(x)}$$

Stabilitatea numerică exprimă mărimea erorilor numerice introduse de algoritm, în ipoteza unor date de intrare exacte, $\| \boldsymbol{P(x)} - \boldsymbol{\tilde{P}(x)} \|$ sau $\dfrac{\| \boldsymbol{P(x)} - \boldsymbol{\tilde{P}(x)} \|}{\| \boldsymbol{P(x)} \|}$.

O eroare relativă de ordinul erorii de rotunjire caracterizează un ***algoritm numeric stabil***.

Un **algoritm numeric stabil** aplicat unei **probleme bine condiţionate** conduce la **rezultate cu precizie foarte bună**.

Un algoritm $\boldsymbol{\tilde{P}}$ destinat rezolvării problemei $\boldsymbol{P}$ este numeric stabil dacă este îndeplinită una din condiţiile:

1. $\tilde{P}(x) \approx P(x)$ pentru orice intrare x;

2. există $\tilde{x}$ apropiat de x, astfel ca $\tilde{P}(x) \approx P(\tilde{x})$

x = datele exacte,

$P(x)$ = soluţia exactă folosind date exacte,

$\tilde{P}(x)$ = soluţia „*calculată*" folosind algoritmul $\tilde{P}$ cu date exacte de intrare

Rezolvarea sistemelor liniare

Istoric

- 1900 î.Hr., Babilon - apar primele probleme legate de ecuaţii liniare simultane
- 300 î.Hr. Babilon - tăbliţă cu următoarea problemă:”*Avem două câmpuri de arie totală 1800 ha. Producţia la hectar pe primul câmp este de 2/3 buşel (=36,3l) iar pe al doilea este de 1/2 buşel. Dacă producţia totală este de 1100 buşeli, să se determine aria fiecărui teren în parte.*

- 200-100 î.Hr. China – *9 capitole despre arta matematică* – metodă de rezolvare foarte asemănatoare eliminării Gauss (*„Avem 3 tipuri de grâu. Ştim că 3 baloturi din primul tip, 2 baloturi din al doilea tip şi 1 balot din al treilea tip cântăresc 39 măsuri. Deasemenea, 2 baloturi din primul tip, 3 baloturi din al doilea tip şi 1 balot din al treilea tip cântăresc 34 măsuri şi 1 balot din primul tip, 2 baloturi din al doilea tip şi 3 baloturi din al treilea tip cântăresc 26 măsuri. Câte măsuri cântăreşte un balot din fiecare tip de grâu”)*

- 1545, Cardan – în *Ars Magna*, propune o regulă (*regula de modo)* pentru rezolvarea unui sistem de 2 ecuaţii cu 2 necunoscute (seamănă cu regula lui Cramer)
- 1683, Seki Kowa, Japonia - ideea de „*determinant*”- „*Method of solving the dissimulated problems”*. Calculează ceea ce astăzi cunoaştem sub numele de determinant, determinanţii matricilor ***2x2***, ***3x3***, ***4x4***, ***5x5*** în legătură cu rezolvarea unor ecuaţii dar nu a sistemelor de ecuaţii.

- 1683, Leibniz într-o scrisoare către l'Hôpital explică faptul că sistemul de ecuaţii:

$$10+11x+12y=0$$
$$20+21x+22y=0$$
$$30+31x+32y=0$$

are soluţie deoarece :

10*21*32+11*22*30+12*20*31=10*22*31+11*20*32+12*21*30

(condiţia ca determinantul matricii coeficienţilor este 0).

Leibniz era convins că o notaţie matematică bună este cheia progresului şi experimentează mai mult de 50 de moduri diferite de a scrie coeficienţii unui sistem de ecuaţii. Leibniz foloseşte termenul de „*rezultant”* în loc de determinant şi a demonstrat regula lui Cramer pentru „rezultanţi”. Ştia că orice determinant poate fi dezvoltat în raport cu o coloană – operaţia se numeşte azi dezvoltarea Laplace.

- 1750, Cramer prezintă o formulă bazată pe determinanţi pentru rezolvarea unui sistem de ecuaţii liniare – *regula lui Cramer* – *„Introduction in the analysis of algebraic curves”* (dă o regulă generală pentru sisteme *n x n:*

  *‚One finds the value of each unknown by forming **n** fractions of which the common denominator has as many terms as there are permutations of **n** things’*

- 1764 Bezout, 1771 Vandermonde, 1772 Laplace – reguli de calcul al determinanţilor

- 1773 Lagrange – prima utilizare implicită a matricilor în legătură cu formele biliniare ce apar la optimizarea unei funcţii reale de 2 sau mai multe variabile (dorea să caracterizeze punctele de maxim şi minim a funcţiilor de mai multe variabile)

- 1800-1801, Gauss introduce noţiunea de „*determinant”* (determină proprietăţile formei pătratice) – *Disquisitiones arithmeticae(1801)*; descrie operaţiile de înmulţire matricială şi inversă a unei matrici în contextul tabloului coeficienţilor unei forme pătratice. Gauss dezvoltă *eliminarea Gaussiană* pe când studia orbita asteroidului Pallas de unde obţine un sistem liniar cu 6 ecuaţii cu 6 necunoscute.

- 1812, Cauchy foloseşte termenul de „*determinant*” în sensul cunoscut azi.
- 1826, Cauchy găseşte valorile proprii şi deduce rezultate legate de diagonalizarea unei matrici. Introduce noţiunea de matrici asemenea şi demonstrează ca acestea au aceeaşi ecuaţie caracteristică. Demonstrează că orice matrice reală simetrică este diagonalizabilă.

- 1850, Sylvester introduce pentru prima data termenul de ***matrice*** (din latină, „uter” – un loc unde ceva se formează sau este produs, „*an oblong arrangement of terms*”)
- 1855, Cayley – algebră matricială, prima definiţie abstarctă a unei matrici. Studiază transformările liniare şi compunerea lor ceea ce îl conduce la operaţiile cu matrici (adunare, înmulţire, înmulţirea cu un scalar, inversa)

- 1858, Cayley în *Memoriu asupra teoriei matricilor : „Sunt multe lucruri de spus despre această teorie a matricilor şi, după părerea mea, această teorie ar trebui să preceadă teoria determinanţilor”*
- Jordan (1870 – Treatise on substitutions and algebraic equations – forma canonică Jordan), Frobenius (1878 – On linear substituions and bilinear forms, rangul unei matrici), Peano

- 1890, Weierstrass – On determinant theory, definiţia axiomatică a determinantului
- 1925, Heisenberg reinventează algebra matricială pentru mecanica cuantică
- 1947, vonNeuman & Goldstine introduc numerele de condiţionare atunci când analizează erorile de rotunjire
- 1948, Turing introduce descompunerea LU a unei matrici
- 1958, Wilkinson dezvoltă factorizarea QR

....

Evaluarea erorii în rezolvarea sistemelor liniare

(condiţionarea sistemelor liniare)

Fie $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b} \in \mathbb{R}^{n}$, $\boldsymbol{x} \in \mathbb{R}^{n}$ şi sist. de ec. liniare:

$$\boldsymbol{Ax = b}$$

$$\boldsymbol{A} \textbf{ nesingulară } \Leftrightarrow \textbf{ det } \boldsymbol{A \neq 0} \Rightarrow \exists \textbf{ sol. sist. } \boldsymbol{x = A^{-1}b}$$

Pentru erorile în datele de intrare facem notaţiile:

- $\boldsymbol{\Delta A} \in \mathbb{R}^{n\times n}$ **eroarea absolută pentru** $\boldsymbol{A}$;
- $\boldsymbol{\Delta b} \in \mathbb{R}^{n}$ **eroarea absolută pentru** $\boldsymbol{b}$;

În realitate se rezolvă sistemul:

$$\left(\boldsymbol{A}+\Delta\boldsymbol{A}\right)\tilde{\boldsymbol{x}}=\boldsymbol{b}+\Delta\boldsymbol{b}$$

soluţia fiind $\tilde{\boldsymbol{x}}$:

$$\tilde{\boldsymbol{x}}=\boldsymbol{x}+\Delta\boldsymbol{x}$$

În mod natural se ridică următoarele probleme :

1. Dacă $\boldsymbol{A}$ este matrice nesingulară, $\Delta\boldsymbol{A}=\boldsymbol{?}$ a.î. $\boldsymbol{A}+\Delta\boldsymbol{A}$ să fie nesingulară ?
2. Pp. $\boldsymbol{A}$ şi $\boldsymbol{A}+\Delta\boldsymbol{A}$ nesingulare care sunt relaţiile între $\frac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}$, $\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}$ şi $\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$?

1. Pp. $\boldsymbol{A}$ nesingulară.

$$\boldsymbol{A+\Delta A = A\left(I_n + A^{-1}\Delta A\right) \rightarrow}$$

$$\boldsymbol{A+\Delta A \text{ nesingulară } \Leftrightarrow \left(I_n + A^{-1}\Delta A\right) \text{ nesingulară}}$$

Propoziţia 5

Fie $\boldsymbol{A}$ nesingulară şi $\|\Delta A\| < \dfrac{\mathbf{1}}{\|A^{-1}\|}$. Atunci $\boldsymbol{I+A^{-1}\Delta A}$ este nesingulară şi avem:

$$\boldsymbol{\left\|\left(I_n + A^{-1}\Delta A\right)^{-1}\right\| \leq \frac{1}{1-\|A^{-1}\|\cdot\|\Delta A\|}}$$

Demonstraţie. Avem:

$$\|\Delta A\| < \frac{1}{\|A^{-1}\|} \Rightarrow \|A^{-1}\Delta A\| \le \|A^{-1}\| \cdot \|\Delta A\| < 1 \overset{\text{Pr.4}}{\Rightarrow} \exists\left(I + A^{-1}\Delta A\right)^{-1}$$

$$\left\|\left(I + A^{-1}\Delta A\right)^{-1}\right\| \le \frac{1}{1 - \|A^{-1}\Delta A\|} \le \frac{1}{1 - \|A^{-1}\| \cdot \|\Delta A\|} .$$

Pp. că A este nesingulară şi $\|\Delta A\| < \dfrac{1}{\|A^{-1}\|}$.

$$(A+\Delta A)(x+\Delta x)=b+\Delta b \Rightarrow (A+\Delta A)\Delta x + Ax + (\Delta A)x = b+\Delta b \Rightarrow$$

$$A\left(I+A^{-1}\Delta A\right)\Delta x = \Delta b - (\Delta A)x \Rightarrow \Delta x = \left(I+A^{-1}\Delta A\right)^{-1} A^{-1}\left[\Delta b - (\Delta A)x\right] \Rightarrow$$

$$\|\Delta x\| \le \left\|\left(I+A^{-1}\Delta A\right)^{-1}\right\| \left\|A^{-1}\right\| \left(\|\Delta b\| + \|\Delta A\|\|x\|\right) \Rightarrow$$

$$\frac{\|\Delta x\|}{\|x\|} \le \frac{\left\|A^{-1}\right\|}{1-\left\|A^{-1}\right\|\|\Delta A\|}\left(\frac{\|\Delta b\|}{\|x\|} + \|\Delta A\|\right) \qquad \textbf{(1)}$$

Din $Ax = b$ obţinem $\|b\| \le \|A\|\|x\| \Rightarrow \dfrac{1}{\|x\|} \le \dfrac{\|A\|}{\|b\|}$ şi ţinând seamă de acest rezultat, din (1) deducem:

$$\frac{\|\Delta x\|}{\|x\|} \le \frac{\left\|A^{-1}\right\|\|A\|}{1-\left\|A^{-1}\right\|\|\Delta A\|}\left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|}\right).$$

k(A) = $\|A^{-1}\|\,\|A\|$ ***numărul de condiţionare*** al matricii ***A***.

Propoziţia 6

Dacă matricea ***A*** este nesingulară şi $\|\Delta A\| < \dfrac{1}{\|A^{-1}\|}$ atunci:

$$\frac{\|\Delta x\|}{\|x\|} \le \frac{k(A)}{1 - k(A)\cdot \dfrac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \right).$$

Din $I_n = A\,A^{-1}$ rezultă $1 = \|I_n\| \le \|A\|\|A^{-1}\| = k(A)$.

$k(A) \geq 1,\ \forall A$ dar dep. de norma matricială naturală utilizată.

O matrice ***A*** pentru care numărul de condiţionare este mare se numeşte matrice ***prost condiţionată (k(A) ‚mare')*****.**

Ax=b* cu *k(A) mare $\rightarrow$ $\dfrac{\|\Delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ poate fi mare chiar dacă erorile relative $\dfrac{\|\Delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}$ **şi** $\dfrac{\|\Delta \boldsymbol{A}\|}{\|\boldsymbol{A}\|}$ sunt mici.

Fie $\boldsymbol{A}$ o matrice simetrică $\boldsymbol{A} = \boldsymbol{A}^{\boldsymbol{T}}$, nesingulară. Utilizând norma matricială subordonată normei vectoriale euclidiene:

$$\|A\|_2 = \sqrt{\rho\left(A^T A\right)} = \sqrt{\rho\left(A^2\right)}$$

$$k(A) = \|A\|_2 \cdot \left\|A^{-1}\right\|_2$$

Matricea simetrică $\boldsymbol{A}$ are valorile proprii reale $\lambda_1, \lambda_2, \ldots, \lambda_n$,

$\boldsymbol{A^2}$ are valorile proprii $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$

$\boldsymbol{A^{-1}}$ are valorile proprii $\dfrac{1}{\lambda_1}, \dfrac{1}{\lambda_2}, \ldots, \dfrac{1}{\lambda_n}$.

$$\left|\lambda_1\right| \le \left|\lambda_2\right| \le .... \le \left|\lambda_n\right| \Rightarrow \rho(A) = \left|\lambda_n\right| \quad \text{şi} \quad \rho\left(A^{-1}\right) = \frac{1}{\left|\lambda_1\right|}$$

$$A = A^T \to \left\|A\right\|_2 = \rho(A) = \left|\lambda_n\right|, \left\|A^{-1}\right\|_2 = \rho\left(A^{-1}\right) = \frac{1}{\left|\lambda_1\right|},$$

$$k_2(A) = \| A \|_2 \cdot \| A^{-1} \|_2 = \frac{\left|\lambda_n\right|}{\left|\lambda_1\right|}$$ ***număr de condiţionare spectral***.

A matrice ortogonală $\to$ ***$k_2(A)=1$***

$$A^T A = A \cdot A^T = I_n \Rightarrow A^{-1} = A^T$$

$$\left\|A\right\|_2 = \sqrt{\rho\left(A^T A\right)} = \sqrt{\rho(I)} = 1 = \left\|A^T\right\|_2$$

$$k_2(A) = \left\|A\right\|_2 \cdot \left\|A^{-1}\right\|_2 = \left\|A\right\|_2 \cdot \left\|A^T\right\|_2 = 1,$$

Matrice aproape singulară dar cu număr de condiţionare mic

$$\boldsymbol{A = \mathrm{diag}\,[1, 0.1, 0.1, ..., 0.1] \in \mathbb{R}^{100\times100} \Rightarrow \det A = 1\cdot(0.1)^{99} = 10^{-99}}$$

$$\boldsymbol{\| A \|_2 = 1\ ,\ \| A^{-1} \|_2 = 10 \ \Rightarrow\ k_2(A) = \| A \|_2\ \| A^{-1} \|_2 = 10}$$

Matrice foarte prost condiţionată cu det. nenul (**det** $\boldsymbol{A}$**=1**)

$$\boldsymbol{A}=\begin{pmatrix} 1 & 2 & 0 & \cdots & 0 \\ 0 & 1 & 2 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 2 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

$$\boldsymbol{A}^{-1}=\begin{pmatrix} 1 & -2 & 4 & \cdots & (-2)^{i-1} & \cdots & (-2)^{n-1} \\ 0 & 1 & -2 & \cdots & (-2)^{i-2} & \cdots & (-2)^{n-2} \\ \vdots & & & & & & \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

$$\|A\|_\infty = \|A\|_1 = 3 \quad ,$$

$$\|A^{-1}\|_\infty = \|A^{-1}\|_1 = 1 + 2 + \cdots + 2^{n-1} = 2^n - 1$$

$$n = 100 \;\Rightarrow\; k(A) = \|A\|_\infty \|A^{-1}\|_\infty = \|A\|_1 \|A^{-1}\|_1 = 3 \cdot (2^{100} - 1)$$

$$\det A = 1$$

$$\begin{cases} x + y = 2 \\ x + 1.001y = 2 \end{cases}, \quad \begin{cases} x + y = 2 \\ x + 1.001y = 2.001 \end{cases} \quad k(A) = 4002$$

$$x = 2\ , \ \ y = 0 \qquad\qquad x = 1\ , \ \ y = 1$$

$$\begin{cases} 400x - 201y = 200 \\ -800x + 401y = -200 \end{cases}, \quad \begin{cases} 401x - 201y = 200 \\ -800x + 401y = -200 \end{cases} \quad k(A) = 4002$$

$$x = -100\ , \ \ y = -200 \qquad\qquad x = 40000\ , \ \ y = 79800$$

$$\begin{cases} 1.2969x + 0.8648\,y = 0.8642 \\ 0.2161x + 0.1441\,y = 0.1440 \end{cases} \quad x = 2\,,\ y = -2 \qquad k_2(A) = 249730000$$

$$\bar{x} = 0.9911,\ \bar{y} = -0.4870\ ,$$

$$r = b - Az = \begin{pmatrix} 0.8642 \\ 0.1440 \end{pmatrix} - \begin{pmatrix} 1.2969 & 0.8648 \\ 0.1441 & 0.1441 \end{pmatrix} \begin{pmatrix} 0.9911 \\ -0.4870 \end{pmatrix} = \begin{pmatrix} 10^{-8} \\ -10^{-8} \end{pmatrix}$$

Matricea Hilbert

$$H = (h_{ij})_{i,j=1}^{n} \quad , \quad h_{ij} = \frac{1}{i+j-1} = \int_0^1 x^{i+j-2}\,dx$$

$$k_2(H_n) \approx \frac{(\sqrt{2}+1)^{4(n+1)}}{2^{\frac{15}{4}}\sqrt{\pi n}} \sim e^{3.5n}$$

n	$k_2(H_n)$	n	$k_2(H_n)$
1	1	7	$4.753\cdot10^{8}$
2	19.281	8	$1.526\cdot10^{10}$
3	$5.241\cdot10^{2}$	9	$4.932\cdot10^{11}$
4	$1.551\cdot10^{4}$	10	$1.602\cdot10^{13}$
5	$4.766\cdot10^{5}$	11	$5.220\cdot10^{14}$
6	$1.495\cdot10^{7}$	12	$1.678\cdot10^{16}$

$$H^{-1}=(g_{ij})\quad g_{ij}=\frac{(-1)^{i+j}}{(i+j-1)}\frac{(n+i-1)!\,(n+j-1)!}{\left[(i-1)!(j-1)!\right]^{2}(n-i)!(n-j)!}$$

Metode numerice de rezolvarea sistemelor liniare

Fie matricea nesingulară $A \in \mathbb{R}^{n \times n}$ şi $b \in \mathbb{R}^n$. Rezolvarea sistemului de ecuaţii liniare $\boldsymbol{Ax=b}$ se poate face folosind ***regula lui Cramer:***

$$x_i = \frac{\det A_i(b)}{\det A}, i = 1, \ldots, n,$$

în care $A_i(b)$ se obţine din matricea A prin înlocuirea coloanei i cu vectorul b.

Algoritmul dat de regula lui Cramer este foarte costisitor din punct de vedere al resurselor şi instabil numeric.

Din aceste motive s-au căutat alte metode de aproximare a soluţiei $\boldsymbol{x}$. Unul din cele mai folosiţi algoritmi este **algoritmul de eliminare Gauss :**

$$\boldsymbol{Ax=b} \;\leftrightarrow\; \boldsymbol{\tilde{A}x = \tilde{b}} \textbf{ cu } \tilde{A} \text{ matrice superior triunghiulară}$$

$$\boldsymbol{x = A^{-1}b = \tilde{A}^{-1}\tilde{b}} \quad \textbf{(notăm } \boldsymbol{Ax = b \sim \widetilde{A}x = \tilde{b}}\textbf{)}$$

Metoda substituţiei

Fie sistemul liniar $\boldsymbol{Ax = b}$ unde matricea sistemului $\boldsymbol{A}$ este triunghiulară. Pentru a găsi soluţia unică a sistemului, trebuie ca matricea să fie nesingulară. Determinantul matricilor triunghiulare este dat de formula:

$$\det A = a_{11} a_{22} \cdots a_{nn}$$

$$\det A \neq 0 \;\; , a_{ii} \neq 0 \quad \forall i = 1, 2, \ldots, n$$

Vom considera întâi cazul când matricea A este inferior triunghiulară. Sistemul are forma:

$$\begin{aligned}
&a_{11}x_1 = b_1 \\
&a_{21}x_1 + a_{22}x_2 = b_2 \\
&\vdots \\
&a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ii}x_i = b_i \\
&\vdots \\
&a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{ni}x_i + \cdots + a_{nn}x_n = b_i
\end{aligned} \qquad (1)$$

Necunoscutele $x_1, x_2, \ldots, x_n$, se deduc folosind ecuaţiile sistemului de la prima către ultima.

Din prima ecuaţie se deduce x_1:

$$x_1 = \frac{b_1}{a_{11}} \qquad (2)$$

Din a doua ecuaţie , utilizând valoarea x_1 din (2), obţinem x_2:

$$x_2 = \frac{b_2 - a_{21}x_1}{a_{22}}$$

Când ajungem la ecuaţia $\boldsymbol{i}$:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ii-1}x_{i-1} + a_{ii}x_i = b_i$$

folosind variabilele $x_1, x_2, \ldots, x_{i-1}$ calculate anterior, avem:

$$x_i = \frac{b_i - a_{i1}x_1 - \cdots - a_{ii-1}x_{i-1}}{a_{ii}}$$

Din ultima ecuaţie se deduce x_n astfel:

$$x_n = \frac{b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{nn-1}x_{n-1}}{a_{nn}}$$

Algoritmul de calcul al soluţiei sistemelor (1) cu matrice inferior triunghiulară este următorul:

$$\boldsymbol{x_i = \frac{b_i - \sum_{j=1}^{i-1} a_{ij} x_j}{a_{ii}}, i = 1, 2, \ldots, n-1, n}$$

Acest algoritm se numeşte ***metoda substituţiei directe***.

Vom considera, în continuare sistemul (1) cu matrice superior triunghiulară :

$$\begin{array}{l}
a_{11}x_1 + \cdots + a_{1i}x_i + \cdots + a_{1n-1}x_{n-1} + a_{1n}x_n = b_1 \\
\ddots \\
\qquad\qquad a_{ii}x_i + \cdots + a_{in-1}x_{n-1} + a_{in}x_n = b_i \\
\qquad\qquad \ddots \\
\qquad\qquad\qquad a_{n-1n-1}x_{n-1} + a_{n-1n}x_n = b_{n-1} \\
\qquad\qquad\qquad\qquad\qquad a_{nn}x_n = b_n
\end{array}$$

Necunoscutele $x_1, x_2, \ldots, x_n$ se deduc pe rând, folosind ecuaţiile sistemului, de la ultima către prima.

Din ultima ecuaţie găsim x_n:

$$x_n = \frac{b_n}{a_{nn}}$$

Folosind valoarea lui x_n dedusă mai sus, din penultima ecuaţie obţinem:

$$x_{n-1} = \frac{b_{n-1} - a_{n-1n} x_n}{a_{n-1n-1}}$$

Când ajungem la ecuaţia $\boldsymbol{i}$:

$$a_{ii}x_i + a_{ii+1}x_{i+1} + \cdots + a_{in}x_n = b_i$$

se cunosc deja $x_{i+1}, x_{i+2}, \ldots, x_n$ de unde ducem:

$$x_i = \frac{b_i - a_{ii+1}x_{i+1} - \cdots - a_{in}x_n}{a_{ii}}$$

Din prima ecuaţie găsim valoarea lui x_1:

$$x_1 = \frac{b_1 - a_{12}x_2 - \cdots - a_{1n}x_n}{a_{11}}$$

Procedeul descris mai sus se numeşte de ***metoda substituţiei inverse*** pentru rezolvarea sistemelor liniare cu matrice superior triunghiulară:

$$x_i = \frac{b_i - \sum_{j=i+1}^{n} a_{ij} x_j}{a_{ii}}, i = n, n-1, \cdots, 2, 1.$$

M - numărul de operaţii ***, /** (înmulţiri/împărţiri) efectuate

A - numărul operaţiilor $\pm$ (adunări/scăderi) efectuate.

Atunci pentru calculul componentei x_i se efectuează

$M=n-i+1$, $A=n-i$ şi în total:

$$M=\sum_{i=n}^{1}(n-i+1)=\sum_{k=1}^{n}k=\frac{n(n+1)}{2},$$

$$A=\sum_{i=n}^{1}(n-i)=\sum_{k=1}^{n-1}k=\frac{n(n-1)}{2}$$

Efortul de calcul pentru metoda substituţiei directe este

$$M=\frac{n(n+1)}{2} \qquad A=\frac{n(n-1)}{2}.$$

Algoritmul de eliminare Gauss

Algoritmul se realizează în ***n-1*** paşi prin transformarea sistemului dat într-un sistem echivalent cu matrice triunghiulară superior.

Pas 1

la acest pas se obţine sistemul:

$\boldsymbol{A}^{(1)}\boldsymbol{x} = \boldsymbol{b}^{(1)} \sim \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$ unde $\boldsymbol{A}^{(1)}$ are prima coloană în formă superior triunghiulară.

Pas 2

se construieşte sistemul

$\boldsymbol{A^{(2)}x = b^{(2)} \sim Ax = b,}$ unde $\boldsymbol{A^{(2)}}$ are primele două coloane în formă superior triunghiulară.

$\vdots$

Pasul r

se obţine sistemul $\boldsymbol{A^{(r)}x = b^{(r)} \sim Ax = b,}$ unde $\boldsymbol{A^{(r)}}$ are primele ***r*** coloane în formă superior triunghiulară.

$\vdots$

Pasul n-1 :

se obţine sistemul

$\boldsymbol{A^{(n-1)}x = b^{(n-1)} \sim Ax = b,}$ unde $\boldsymbol{A^{(n-1)}}$ are primele ***n-1*** coloane în formă superior triunghiulară.

Dacă la un anumit pas matricea $A^{(r)}$ **nu poate fi construită** aceasta ne va arăta că matricea A **este singulară**.

În realizarea acestor paşi se utilizează următoarele operaţii elementare:

- **înmulţirea unei ecuaţii cu un factor şi adunarea la altă ecuaţie**;
- **interschimbarea a două linii şi/sau două coloane în matricea** A.

Pasul 1

Intrare : sistemul $\boldsymbol{Ax=b}$
Ieşire : sistemul $\boldsymbol{A^{(1)}x = b^{(1)} \sim Ax = b}$, matr $\boldsymbol{A^{(1)}}$ are prima coloană în formă superior triunghiulară.

Fie ecuaţia $\boldsymbol{i}$, cu $\boldsymbol{i=1,\ldots,n}$

$$\boldsymbol{E_i} \ \boldsymbol{:} \qquad \boldsymbol{a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i}.$$

Presupunem $\boldsymbol{a_{11} \neq 0}$. Operaţiile efectuate au ca obiectiv anularea coeficienţilor lui $\boldsymbol{x_1}$ din ecuaţiile de la $\boldsymbol{2}$ la $\boldsymbol{n}$ şi sunt descrise în continuare:

$$E_1 * \left(\frac{-a_{21}}{a_{11}} \right) + E_2 = E_2^{(1)} \quad \Rightarrow \quad a_{21}^{(1)} = 0$$

$$\vdots$$

$$E_1 * \left(\frac{-a_{i1}}{a_{11}} \right) + E_i = E_i^{(1)} \quad \Rightarrow \quad a_{i1}^{(1)} = 0$$

$$\vdots$$

$$E_1 * \left(\frac{-a_{n1}}{a_{11}} \right) + E_n = E_n^{(1)} \quad \Rightarrow \quad a_{n1}^{(1)} = 0$$

Sistemul obţinut prin aceste operaţii are forma:

$$\begin{cases} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \cdots\cdots\cdots + a_{1n}^{(1)}x_n = b_1^{(1)} \\ \qquad\qquad a_{22}^{(1)}x_2 + \cdots\cdots\cdots + a_{2n}^{(1)}x_n = b_2^{(1)} \\ \qquad\qquad \vdots \\ \qquad\qquad a_{i2}^{(1)}x_2 + \cdots\cdots\cdots + a_{in}^{(1)}x_n = b_i^{(1)} \\ \qquad\qquad \vdots \\ \qquad\qquad a_{n2}^{(1)}x_2 + \cdots\cdots\cdots + a_{nn}^{(1)}x_n = b_n^{(1)} \end{cases}$$

Pas 2

Intrare : $\boldsymbol{A^{(1)}x = b^{(1)}}$

Ieşire : $\boldsymbol{A^{(2)}x = b^{(2)} \sim Ax = b,\ A^{(2)}}$ are primele două coloane în formă superior triunghiulară.

Se presupune $\boldsymbol{a_{22}^{(1)} \neq 0}$ şi se urmăreşte anularea elementelor $\boldsymbol{a_{32}^{(2)}, a_{42}^{(2)}, \ldots, a_{n2}^{(2)}}$ (transformarea coloanei 2 în formă superior triunghiulară). Operaţiile efectuate asupra ecuaţiilor $\boldsymbol{E_i^{(1)}, i = 3, \cdots, n}$ sunt următoarele :

$$
\begin{cases}
E_2^{(1)} * \left(-\dfrac{a_{32}^{(1)}}{a_{22}^{(1)}}\right) + E_3^{(1)} = E_3^{(2)} \quad \Rightarrow \quad a_{32}^{(2)} = 0; \\
\vdots \\
E_2^{(1)} * \left(-\dfrac{a_{i2}^{(1)}}{a_{22}^{(1)}}\right) + E_i^{(1)} = E_i^{(2)} \quad \Rightarrow \quad a_{i2}^{(2)} = 0; \\
\vdots \\
E_2^{(1)} * \left(-\dfrac{a_{n2}^{(1)}}{a_{22}^{(1)}}\right) + E_n^{(1)} = E_n^{(2)} \quad \Rightarrow \quad a_{n2}^{(2)} = 0;
\end{cases}
$$

Se observă că nu se schimbă forma superior triunghiulară a primei coloane.

Pas r

Intrare : $\boldsymbol{A}^{(r-1)}\boldsymbol{x} = \boldsymbol{b}^{(r-1)}$

Ieşire : $\boldsymbol{A}^{(r)}\boldsymbol{x} = \boldsymbol{b}^{(r)} \sim \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{A}^{(r)}$ are primele $\boldsymbol{r}$ coloane în formă superior triunghiulară.

Sistemul are forma următoare:

$$
\begin{cases}
a_{11}^{(r-1)}x_1 + \cdots + a_{1r}^{(r-1)}x_r + \cdots + a_{1n}^{(r-1)}x_n = b_1^{(r-1)} \\
\quad \ddots \\
\qquad \ddots \\
\qquad\qquad a_{rr}^{(r-1)}x_r + \cdots + a_{rn}^{(r-1)}x_n = b_r^{(r-1)} \\
\qquad\qquad a_{r+1r}^{(r-1)}x_r + \cdots + a_{r+1n}^{(r-1)}x_n = b_{r+1}^{(r-1)} \\
\qquad\qquad \vdots \\
\qquad\qquad a_{ir}^{(r-1)}x_r + \cdots + a_{in}^{(r-1)}x_n = b_i^{(r-1)} \\
\qquad\qquad \vdots \\
\qquad\qquad a_{nr}^{(r-1)}x_r + \cdots + a_{nn}^{(r-1)}x_n = b_n^{(r-1)}
\end{cases}
$$

Presupunem $a_{rr}^{(r-1)} \neq 0$.

Vom urmări anularea elementelor $a_{r+1r}^{(r)}, a_{r+2r}^{(r)}, \cdots, a_{nr}^{(r)}$.

$$\begin{cases} E_r^{(r-1)} * \left(-\dfrac{a_{r+1r}^{(r-1)}}{a_{rr}^{(r-1)}} \right) + E_{r+1}^{(r-1)} = E_{r+1}^{(r)} \quad \Rightarrow \quad a_{r+1r}^{(r)} = 0; \\ \vdots \\ E_r^{(r-1)} * \left(-\dfrac{a_{ir}^{(r-1)}}{a_{rr}^{(r-1)}} \right) + E_i^{(r-1)} = E_i^{(r)} \quad \Rightarrow \quad a_{ir}^{(r)} = 0; \\ \vdots \\ E_r^{(r-1)} * \left(-\dfrac{a_{nr}^{(r-1)}}{a_{rr}^{(r-1)}} \right) + E_n^{(r-1)} = E_n^{(r)} \quad \Rightarrow \quad a_{nr}^{(r)} = 0; \end{cases}$$

Se observă că nu se schimbă forma superior triunghiulară a primelor ***r-1*** coloane.

La fiecare pas s-a făcut ipoteza $a_{rr}^{(r-1)} \neq 0$. Elementul $a_{rr}^{(r-1)}$ poartă numele de ***pivot***. În cazul în care elementul pivot este nul se pot aplica următoarele strategii, numite de ***pivotare***:

Pivotare ($a_{rr}^{(r-1)} \neq 0$?)

1^0 ***Fără pivotare***

Se caută primul indice $i_0 \in \{r, r+1, \cdots, n\}$ astfel încât $a_{i_0 r}^{(r-1)} \neq 0$. Se interschimbă liniile i_0 şi r.

Să observăm că în procesul de calcul la pasul r intervine factorul $\dfrac{1}{a_{rr}^{(r-1)}}$ astfel că valori mici ale lui $\left|a_{rr}^{(r-1)}\right|$ conduc la amplificarea erorilor de calcul. Pentru a asigura stabilitatea numerică a procesului de calcul este de dorit ca $\left|a_{rr}^{(r-1)}\right|$ să fie ***'mare'***.

2^0 ***Pivotare parţială***

Se determină indicele $\boldsymbol{i_0}$:

$$\left|a_{i_0 r}^{(r-1)}\right| = \max\left\{\left|a_{ir}^{(r-1)}\right|; i = r,\dots,n\right\}$$

şi se interschimbă liniile $\boldsymbol{i_0}$, $\boldsymbol{r}$ dacă $\boldsymbol{i_0 \neq r}$.

3^0 ***Pivotare totală***

Se determină indicii $\boldsymbol{i_0}$ şi $\boldsymbol{j_0}$:

$$\left|a_{i_0 j_0}^{(r-1)}\right| = \max\left\{\left|a_{ij}^{(r-1)}\right|; i = r,\dots,n\,, j = r,\dots,n\right\}$$

şi se interschimbă liniile $\boldsymbol{i_0}$, $\boldsymbol{r}$ dacă $\boldsymbol{i_0 \neq r}$ şi coloanele $\boldsymbol{j_0}$, $\boldsymbol{r}$ dacă $\boldsymbol{j_0 \neq r}$

Schimbarea coloanelor implică schimbarea ordinii variabilelor astfel încât în final va trebui refăcută ordinea iniţială a variabilelor.

Dacă după pivotare elementul pivot rămâne nul, $\boldsymbol{a_{rr}^{(r-1)} = 0}$**,** atunci putem deduce că $\boldsymbol{A^{(r-1)}}$ este singulară.

În adevăr, dacă în procesul de pivotare parţială

$\boldsymbol{a_{rr}^{(r-1)} = 0}$**,** atunci

$$A^{(r-1)} = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots\cdots & & a_{1n} \\ & \ddots & & & \\ & & a_{rr} = 0 & \cdots & a_{rn} \\ & & 0 & & \\ & & 0 & & \\ & & 0 & \cdots\cdots\cdots & a_{nn} \end{bmatrix} \Rightarrow$$

$$\det A^{(r-1)} = a_{11}^{(r-1)} a_{22}^{(r-1)} \cdots a_{r-1r-1}^{(r-1)} \det \begin{bmatrix} 0 \cdots\cdots a_{rn} \\ 0 \\ \vdots \\ 0 \cdots\cdots a_{nn} \end{bmatrix} = 0$$

Deoarece operaţiile efectuate (cele de inetrschimbare de linii şi/sau coloane) nu au schimbat decât semnul determinantului avem:

$$\det A = \pm \det A^{(r-1)} = 0 \;\Rightarrow\; \det A = 0$$

prin urmare matricea A iniţială este singulară.

Şi în cazul procesului de pivotare totală dacă $a_{rr}^{(r-1)} = 0$, atunci:

$$A^{(r-1)} = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots\cdots & a_{1n} \\ & \ddots & & \\ & & a_{rr} = 0 \cdots & \mathbf{0} \\ & & 0 & \\ & & 0 & \\ & & 0 \cdots\cdots\cdots & \mathbf{0} \end{bmatrix} \Rightarrow$$

$$\det A^{(r-1)} = a_{11}^{(r-1)} a_{22}^{(r-1)} \cdots a_{r-1\,r-1}^{(r-1)} \det \begin{bmatrix} \mathbf{0} \cdots\cdots \mathbf{0} \\ \mathbf{0} \cdots\cdots \mathbf{0} \\ \vdots \qquad \vdots \\ \mathbf{0} \cdots\cdots \mathbf{0} \end{bmatrix} = \mathbf{0}$$

$\det A = \pm \det A^{(r-1)} = \mathbf{0} \ \Rightarrow\ A$ este matrice singulară.

$r = 1;$

pivotare$(r);$

while $(r \le n - 1$ **şi** $|a_{rr}| > \varepsilon)$

// **Pas r**

$*$ **for** $i = r+1, \ldots, n$

$\bullet\ f = -\dfrac{a_{ir}}{a_{rr}};$

$\bullet$ **for** $j = r+1, \ldots, n$

$a_{ij} = a_{ij} + f * a_{rj};$

$\bullet\ a_{ir} = 0;$

$\bullet\ b_i = b_i + f * b_r;$

$*\ r = r + 1;$

$*$ **pivotare**$(r);$

if $(|a_{rr}| \le \varepsilon)$ **'MATRICE SINGULARA'**

else $\{ A \leftarrow A^{(n-1)},\ b \leftarrow b^{(n-1)}$

se rezolvă sistemul triunghiular superior $Ax = b\}$

Numărul de operaţii efectuate la pasul $\boldsymbol{r}$ şi în total este:

$$(n-r)\left[1M+(n-r)A+(n-r)M+1A+1M\right] \quad \Rightarrow$$

$$\mathbf{M:}\quad \sum_{r=1}^{n-1}(n-r)^2+2\sum_{r=1}^{n-1}(n-r)=\frac{(n-1)n(2n+5)}{6},$$

$$\mathbf{A:}\quad \sum_{r=1}^{n-1}(n-r)^2+\sum_{r=1}^{n-1}(n-r)=\frac{(n-1)n(n+1)}{3},$$

$$\mathbf{M:}\ \frac{n^3}{3}+\mathcal{O}\left(n^2\right) \qquad ; \qquad \mathbf{A:}\ \frac{n^3}{3}+\mathcal{O}\left(n^2\right)$$

Eliminarea *„chinezească”*

200-100 î.Cr. China – *9 capitole despre arta matematică* – metodă de rezolvare foarte asemănatoare eliminării Gauss

„Avem 3 tipuri de grâu. Ştim că 3 baloturi din primul tip, 2 baloturi din al doilea tip şi 1 balot din al treilea tip cântăresc 39 măsuri. Deasemenea, 2 baloturi din primul tip, 3 baloturi din al doilea tip şi 1 balot din al treilea tip cântăresc 34 măsuri şi 1 balot din primul tip, 2 baloturi din al doilea tip şi 3 baloturi din al treilea tip cântăresc 26 măsuri. Câte măsuri cântăreşte un balot din fiecare tip de grâu”

Notaţia actuală:

$$3b_1 + 2b_2 + b_3 = 39$$
$$2b_1 + 3b_2 + b_3 = 34$$
$$b_1 + 2b_2 + 3b_3 = 26$$

Notaţia *chinezească*

1	**2**	**3**
2	**3**	**2**
3	**1**	**1**
26	**34**	**39**

Pasul 1

Se înmulţeşte coloana a doua cu 3 şi se scade din ea coloana a treia atât timp cât este posibil.

Se înmulţeşte prima coloană cu 3 şi se scade din ea coloana a treia atât timp cât este posibil.

Se ajunge la forma:

0	**0**	**3**
4	**5**	**2**
8	**1**	**1**
39	**24**	**39**

Pasul 2

Se înmulţeşte prima coloană cu 5 şi se scade din ea coloana a doua atât timp cât este posibil.

Se ajunge la forma:

0	**0**	**3**
0	**5**	**2**
36	**1**	**1**
99	**24**	**39**

Pentru rezolvare se foloseşte metoda substituţiei inverse pe sistemul obţinut mai sus.