Calcul Numeric

Cursul 3

2020

Anca Ignat

Propoziţia 4

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

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

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

[27.5664062499999982236431605997495353221893310546875,
27.5664062500000017763568394002504646778106689453125).

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

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

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

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

Numerele care apar în calcule și sunt mai mici decât $\boldsymbol{z}$ sunt setate în general la 0 (*underflow*) iar cele mai mari decât $\boldsymbol{Z}$ 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

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

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

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

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

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

Un număr r^* aproximează numărul r cu t cifre exacte dacă t este cel mai mare intreg nenegativ pentru care:

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

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 perturbații temporare,

   - erori de rotunjire: 1/3 , π , 1/7,…

2. Erori de rotunjire în timpul calculelor:

   - datorate capacității limitate de memorare a datelor, operaţiile nu sunt efectuate exact.

3. Erori de <u>discretizare:</u>

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

4. <u>Simplificări</u> în modelul matematic

- idealizări , ignorarea unor parametri.

5. Erori <u>umane</u> 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: $a \neq 0$ $\quad \frac{a - \tilde{a}}{a}$ **sau** $\frac{|a - \tilde{a}|}{|a|}$ **sau** $\frac{\|a - \tilde{a}\|}{\|a\|}$

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

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

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

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

a_1 cu eroare relativă δ_{a_1} şi a_2 cu eroare relativă δ_{a_2} :

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

Condiționare ←→ stabilitate

Condiționarea 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 aproximație 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 $\dfrac{\| x-\tilde{x} \|}{\| x \|}$ 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 ***x***). O problemă este bine-condiţionată dacă este bine-condiționată în orice punct.

Se consideră polinomul Wilkinson:

$$\boldsymbol{w(x) = (x-1)(x-2)\cdots(x-20) = x^{20} - 210x^{19} + P_{18}(x)}$$

Dacă se schimbă coeficientul ***210*** al lui $\boldsymbol{x^{19}}$ cu

$$\boldsymbol{210 - 2^{-23} = -210.0000001192}$$

soluțiile (cu 5 zecimale exacte) noului polinom sunt:

1.00000 2.00000 3.00000 4.00000 5.00000 6.00001 6.99970 8.00727 8.91725 20.84691 $10.09527 \pm 0.64350i$ $11.79363 \pm 1.65233i$ $13.99236 \pm 2.51883i$ $16.73074 \pm 2.81262i$ $19.50244 \pm 1.94033i$

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,

$\|\boldsymbol{P(x) - \tilde{P}(x)}\|$ sau $\dfrac{\|\boldsymbol{P(x) - \tilde{P}(x)}\|}{\|\boldsymbol{P(x)}\|}$.

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

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 asemănătoare 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. De asemenea, 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 matricelor ***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 matricei 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 matricelor 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 matrice 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 găsește valorile proprii şi deduce rezultate legate de diagonalizarea unei matrice. Introduce noțiunea de matrice asemenea şi demonstrează ca acestea au aceeași ecuație 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 abstractă a unei matrice. Studiază transformările liniare şi compunerea lor ceea ce îl conduce la operaţiile cu matrice (adunare, înmulţire, înmulţirea cu un scalar, inversa)

- 1858, Cayley în *Memoriu asupra teoriei matricelor : „Sunt multe lucruri de spus despre această teorie a matricelor 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)
- 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 matrice
- 1958, Wilkinson dezvoltă factorizarea QR

....

Evaluarea erorii în rezolvarea sistemelor liniare

(condiţionarea sistemelor liniare)

Fie $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n\times n}}$, $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ 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 notațiile:

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

În realitate se rezolvă sistemul:

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

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 $\dfrac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}$, $\dfrac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}$ şi $\dfrac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$?

1. Pp. $\boldsymbol{A}$ nesingulară.

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

$$\boldsymbol{A} + \Delta\boldsymbol{A} \textbf{ nesingulară } \Leftrightarrow \left(\boldsymbol{I}_n + \boldsymbol{A}^{-1}\Delta\boldsymbol{A}\right) \textbf{ nesingulară}$$

Propoziţia 5

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

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

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 $\|\Delta A\| < \frac{1}{\|A^{-1}\|}$.

$$(\boldsymbol{A}+\Delta\boldsymbol{A})(\boldsymbol{x}+\Delta\boldsymbol{x})=\boldsymbol{b}+\Delta\boldsymbol{b}\Rightarrow(\boldsymbol{A}+\Delta\boldsymbol{A})\Delta\boldsymbol{x}+\boldsymbol{A}\boldsymbol{x}+(\Delta\boldsymbol{A})\boldsymbol{x}=\boldsymbol{b}+\Delta\boldsymbol{b}\Rightarrow$$

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

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

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

Din $\boldsymbol{Ax}=\boldsymbol{b}$ obţinem $\|\boldsymbol{b}\|\le\|\boldsymbol{A}\|\|\boldsymbol{x}\|\Rightarrow\dfrac{\mathbf{1}}{\|\boldsymbol{x}\|}\le\dfrac{\|\boldsymbol{A}\|}{\|\boldsymbol{b}\|}$ şi ţinând seamă

de acest rezultat, din (1) deducem:

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

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

Propoziţia 6

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

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

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

k(A)* ≥ *1*, ∀*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 $\boldsymbol{A = A^T}$, nesingulară. Utilizând norma matricială subordonată normei vectoriale euclidiene:

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

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

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 $\dfrac{1}{\lambda_1}, \dfrac{1}{\lambda_2}, \ldots, \dfrac{1}{\lambda_n}$.

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

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

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

A matrice ortogonală $\to$ $k_2(A)=1$

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

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

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

Matrice aproape singulară dar cu număr de condiţionare mic

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

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

Matrice foarte prost condiţionată cu det. nenul (**det *A*=1**)

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

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

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

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

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

$$\det A = 1$$

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

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

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

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

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

$$\overline{x} = 0.9911, \; \overline{y} = -0.4870 \quad ,$$

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

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 ecuații liniare ***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 obține 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 și instabil numeric.

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

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

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

C:\ D\ cn info3 2019-2020 sem ii\curs\curs 04\Cramer.exe

```
 n = 11

 Eliminarea Gauss
29-Feb-20 11:52:29 AM
 Timp in milisecunde (eliminare Gauss) 0
29-Feb-20 11:52:29 AM


 Eroarea |Ax-b| = 1.69104133049023E-14


 Eroarea |x-x_exact| = 1.64048256273382E-14

 Solutia exacta : 1,2, ...


 Regula lui Cramer
29-Feb-20 11:52:29 AM
 Timp in milisecunde (Cramer) 93501
29-Feb-20 11:54:02 AM


 Eroarea |Ax-b| = 1.34917541176323E-09


 Eroarea |x-x_exact| = 6.96403558072442E-10
```

C:\ D\ cn info3 2019-2020 sem ii\curs\curs 04\Cramer.exe

```
 n = 12

 Eliminarea Gauss
29-Feb-20 12:26:43 PM
 Timp in milisecunde (eliminare Gauss) 0
29-Feb-20 12:26:43 PM


 Eroarea |Ax-b| = 1.06882473824949E-14


 Eroarea |x-x_exact| = 2.48991194831634E-14

 Solutia exacta : 1,2, ...


 Regula lui Cramer
29-Feb-20 12:26:43 PM
 Timp in milisecunde (Cramer) 1518648
29-Feb-20 12:52:01 PM


 Eroarea |Ax-b| = 2.45286120116051E-11


 Eroarea |x-x_exact| = 2.83016046762119E-11
```

C:\ D\ cn info3 2019-2020 sem ii\curs\curs 04\Cramer.exe

```
n = 13

Eliminarea Gauss
29-Feb-20 1:01:45 PM
Timp in milisecunde (eliminare Gauss) 1
29-Feb-20 1:01:45 PM


Eroarea |Ax-b| = 2.54926160776485E-14


Eroarea |x-x_exact| = 3.74706364093036E-14

Solutia exacta : 1,2, ...


Regula lui Cramer
29-Feb-20 1:01:45 PM
Timp in milisecunde (Cramer) 11442925
29-Feb-20 4:12:28 PM


Eroarea |Ax-b| = 7.9383093591141E-09


Eroarea |x-x_exact| = 6.22578860104344E-09
```

C:\ D\ cn info3 2019-2020 sem ii\curs\curs 04\Cramer.exe

```
 n = 6000

 Eliminarea Gauss
29-Feb-20 4:25:12 PM
 Timp in milisecunde (eliminare Gauss) 891327
29-Feb-20 4:40:03 PM


 Eroarea |Ax-b| = 5.85962827939688E-06


 Eroarea |x-x_exact| = 2.94497871278736E-06

 Solutia exacta : 1,2, ...


 Regula lui Cramer
29-Feb-20 4:40:04 PM
```