Calcul Numeric

Curs 3

Conf. dr. Anca Ignat

Conf. dr. Andreea Arusoaie

February 28, 2026

Norme matriceale naturale (induse)

Fie $\|\cdot\|_v$ și $\|\cdot\|_{v,P}$ norme vectoriale, iar $\|\cdot\|_i$ și respectiv $\|\cdot\|_{i,P}$ normele matriceale induse de acestea.

$$\begin{array}{ccc} \|\cdot\|_v & \longrightarrow & \|\cdot\|_i \\ \downarrow & & ? \\ \|\cdot\|_{v,P} & \longrightarrow & \|\cdot\|_{i,P} \end{array}$$

Norma vectorială ponderată $\|\cdot\|_{v,P}$ este definită prin $\|\mathbf{x}\|_{v,P} = \|P\mathbf{x}\|_v$, unde P este o matrice inversabilă. Astfel, norma matriceală indusă corespunzătoare este

$$\|A\|_{i,P} = \|PAP^{-1}\|_i.$$

În particular, dacă

$$P = \text{diag}(p_1, p_2, \ldots, p_n), p_i > 0, \ \forall i = 1, 2, \ldots, n,$$

atunci avem $PAP^{-1} = \left(\frac{p_i}{p_j} a_{ij}\right)_{i,j=1,\ldots,n}$

Valori și vectori proprii

Definiții. Fie $A \in \mathbb{R}^{n\times n}$ o matrice.

- Se numește *valoare proprie* (sau *autovaloare*) a matricei A un număr complex $\lambda \in \mathbb{C}$ pentru care există un vector nenul $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} \neq 0$, astfel încât

$$A\mathbf{x} = \lambda\mathbf{x}.$$

  Vectorul $\mathbf{x}$ se numește *vector propriu* (sau *autovector*) asociat valorii proprii λ.

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

Valori şi vectori proprii

- Matricea $\lambda I_n - A$ este singulară ($\det(\lambda I_n - A) = 0$).
- Polinomul

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

  se numește *polinom caracteristic* asociat matricei A.
- Gradul polinomului caracteristic este $grad(p_A) = n$, iar acesta are exact n rădăcini care sunt valorile proprii ale matricei A.
- Se numește *rază spectrală* a matricei A numărul

$$\rho(A) = \max_{i=1,\ldots,n} \left\{|\lambda_i|;\ \lambda_i\text{-valorile proprii ale matricei } A\right\}.$$

Valori și vectori proprii

Pentru norma vectorială

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2},$$

norma matriceală indusă este

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

Aceasta se numește *norma spectrală* a matricei A.

Propoziţie

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

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

Numere în format binar

În anul 1985, IEEE a publicat raportul intitulat *Binary Floating Point Arithmetic Standard 754-1985*, urmat de o actualizare în 2008, *IEEE 754-2008*, care furnizează standarde pentru numerele în virgulă mobilă binare și decimale, formate de interschimbare a tipurilor de date, algoritmi de rotunjire aritmetică și tratarea excepțiilor.

Aceste standarde sunt respectate de toți producătorii de calculatoare care folosesc arhitectura în virgulă mobilă.

Numere în format binar

O reprezentare binară pe 64 de biți a unui număr real se face astfel:

- primul bit este *bitul de semn*;
- următorii 11 biți reprezintă *exponentul c*;
- următorii 52 de biți conțin informații despre partea fracționară f, numită și *mantisă*.

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

$$\underbrace{0}_{\text{semn}}\ \underbrace{10000000011}_{\text{exponent}}\ \underbrace{1011100100010000000000000000000000000000000000000000}_{\text{mantisă}}$$

$$= 27.56640625$$

Valoarea reală reprezentată este cuprinsă în intervalul:

$$[\,27.5664062499999982236431605997495353221893310546875,$$
$$27.5664062500000017763568394002504646778106689453125\,).$$

Limitele reprezentării în virgulă mobilă

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

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

iar cel mai mare număr care poate fi reprezentat este notat cu Z, si are $s = 0, c = 2046, f = 1 - 2^{-52}$, adică

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

Numerele care apar în calcule și sunt mai mici decât z sunt, în general, rotunjite la zero, fenomen cunoscut sub numele de *underflow*.

Numerele care depășesc valoarea maximă Z conduc, de obicei, la *overflow*, situație care provoacă oprirea calculelor.

Reprezentări numerice

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

Reprezentarea zecimală

Reprezentarea zecimală folosind k cifre constă în aproximarea unui număr real printr-un număr cu un număr finit de cifre zecimale.

$$\pm 0.d_1 d_2 \dots d_k \times 10^n,\ 1 \le d_1 \le 9,\ 0 \le d_i \le 9,\ i = 2, \dots, k$$

Orice număr real y

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

poate fi reprezentat folosind k cifre printr-o simplă *trunchiere*:

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

Reprezentarea zecimală – Rotunjire

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

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

- Dacă $d_{k+1} \geq 5$, atunci se adaugă 1 la cifra d_k pentru a obține fl(y) (*round up*);
- dacă $d_{k+1} < 5$, atunci se face trunchierea la k cifre (*round down*).

Reprezentarea zecimală – Rotunjire

Aproximare cu cifre exacte

Un număr r^* aproximează numărul real r cu t cifre exacte dacă t este cel mai mare întreg nenegativ pentru care are loc inegalitatea:

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

În cazul *trunchierii* avem:

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

iar când se face *rotunjirea* avem:

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

Reprezentarea zecimală

Operaţii 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 de perturbații temporare;
- erori de rotunjire, datorate reprezentării numerice finite a unor valori reale, de exemplu:

$$\frac{1}{3},\ \pi,\ \frac{1}{7},\ \ldots$$

Surse de erori în calculele numerice

2. Erori de modelare

- *erori de discretizare*:
  - aproximarea limitei unui șir;
  - aproximarea sumei unei serii;
  - aproximarea funcțiilor neliniare prin funcții liniare;
  - aproximarea derivatei unei funcții....
- *simplificări în modelul matematic*:
  - idealizări ale fenomenului studiat;
  - ignorarea unor parametri considerați neglijabili....

Surse de erori în calculele numerice

3. **Erori în timpul calculelor**
   - erori de rotunjire datorate capacității limitate de memorare a datelor; operațiile aritmetice nu sunt efectuate exact;
   - erori ale bibliotecilor software utilizate (*bug-uri*).

4. **Erori umane**
   - erori în datele de intrare;
   - erori în alegerea sau implementarea algoritmilor;
   - interpretarea sau înțelegerea greșită a problemei.

Eroare absolută și eroare relativă

Fie:

- a – valoarea exactă;
- $\tilde{a}$ – valoarea aproximativă.

Eroare absolută - este definită prin:

$$a - \tilde{a} \quad \text{sau} \quad |a - \tilde{a}| \text{ sau} \quad \|a - \tilde{a}\|.$$

De obicei, se notează:

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

unde Δa reprezintă limita superioară a erorii absolute.

Eroare absolută și eroare relativă

Eroare relativă
Pentru $a \neq 0$, eroarea relativă este definită prin:

$$\frac{a - \tilde{a}}{a} \text{ sau } \frac{|a - \tilde{a}|}{|a|} \text{ sau } \frac{\|a - \tilde{a}\|}{\|a\|}.$$

$$\frac{|a - \tilde{a}|}{|a|} \leq \delta_a.$$

În practică, δ_a se exprimă, de regulă, în procente.

Eroare absolută și eroare relativă

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

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

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

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

a_1 cu eroare relativă δ_{a_1} și a_2 cu eroare relativă δ_{a_2}.

Dacă

$$a = a_1 \cdot a_2 \quad \text{sau} \quad a = \frac{a_1}{a_2},$$

rezultă

$$\delta_a \leq \delta_{a_1} + \delta_{a_2}.$$

Condiționare $\longleftrightarrow$ Stabilitate

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

Fie:

- x – datele exacte de intrare;
- $\tilde{x}$ – o aproximație cunoscută a acestora;
- $P(x)$ – soluția exactă a problemei;
- $P(\tilde{x})$ – soluția problemei cu $\tilde{x}$ ca date de intrare.

Se presupune că s-au efectuat calcule exacte la obținerea soluțiilor $P(x)$ și $P(\tilde{x})$.

Condiționarea numerică a unei probleme

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

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

$$k(x) = \frac{\dfrac{\|P(x)-P(\tilde{x})\|}{\|P(x)\|}}{\dfrac{\|x-\tilde{x}\|}{\|x\|}}, \qquad \text{pentru } x \neq 0 \text{ şi } P(x) \neq 0.$$

Condiționarea numerică a unei probleme

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

Condiționarea este o proprietate locală (se evaluează pentru diverse date de intrare x).

O problemă este bine-condiționată dacă este bine-condiționată în orice punct.

$$\begin{aligned}&\text{Eroarea relativă în datele de ieșire}\\ &\qquad\approx\ \text{Număr de condiționare} \times \text{Eroarea relativă în datele de intrare.}\end{aligned}$$

Polinomul Wilkinson

Se consideră polinomul Wilkinson:

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

Dacă se schimbă coeficientul -210 al lui x^{19} cu

$$-210 - 2^{-23} = -210.0000001192,$$

soluțiile (cu 5 zecimale exacte) ale noului polinom sunt:

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

Stabilitate numerică

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

$$\widetilde{P}(x) \neq P(x).$$

Stabilitatea numerică exprimă mărimea erorilor numerice introduse de algoritm, în ipoteza unor date de intrare exacte.

Eroarea algoritmică poate fi măsurată prin:

$$\|P(x) - \widetilde{P}(x)\| \quad \text{sau} \quad \frac{\|P(x) - \widetilde{P}(x)\|}{\|P(x)\|}.$$

Stabilitate numerică

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

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

Stabilitate numerică – Definiție

Un algoritm $\widetilde{P}$ destinat rezolvării problemei P este numeric stabil dacă este îndeplinită una din condițiile:

1. $\widetilde{P}(x) \approx P(x)$ pentru orice intrare x;
2. există $\tilde{x}$ apropiat de x, astfel încât

$$\widetilde{P}(x) \approx P(\tilde{x}).$$

unde:

- x = datele exacte;
- $P(x)$ = soluția exactă folosind date exacte;
- $\widetilde{P}(x)$ = soluția „calculată" folosind algoritmul $\widetilde{P}$ cu date exacte de intrare.

Rezolvarea sistemelor liniare - Istoric

Istoric

- 1900 î.Hr., Babilon – apar primele probleme legate de ecuații liniare simultane;
- 300 î.Hr., Babilon – tăbliță cu următoarea problemă:

„Avem două câmpuri de arie totală 1800 ha. Producția la hectar pe primul câmp este de 2/3 *bușel (=36,3 l), iar pe al doilea este de* 1/2 *bușel. Dacă producția totală este de* 1100 *bușeli, să se determine aria fiecărui teren în parte."*

Rezolvarea sistemelor liniare – Istoric

- 200–100 î.Hr., China – *„Nouă capitole despre arta matematică”* – metodă de rezolvare foarte asemănătoare eliminării Gauss:

  „Avem 3 tipuri de grâu. Știm că 3 baloturi din primul tip, 2 baloturi din al doilea tip și 1 balot din al treilea tip cântăresc 39 măsuri.

  De asemenea, 2 baloturi din primul tip, 3 baloturi din al doilea tip și 1 balot din al treilea tip cântăresc 34 măsuri, iar 1 balot din primul tip, 2 baloturi din al doilea tip și 3 baloturi din al treilea tip cântăresc 26 măsuri.

  Câte măsuri cântărește un balot din fiecare tip de grâu?”

Rezolvarea sistemelor liniare – Istoric

- 1545, Cardan – în *Ars Magna*, propune o regulă (regula de *modo*) pentru rezolvarea unui sistem de 2 ecuații 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 cunoaștem sub numele de determinant, determinanții matricelor 2×2, 3×3, 4×4, 5×5 în legătură cu rezolvarea unor ecuații, dar nu a sistemelor de ecuații.

Rezolvarea sistemelor liniare – Istoric

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

$$\begin{cases} 10x + 11y + 12z = 0, \\ 20x + 21y + 22z = 0, \\ 30x + 31y + 32z = 0, \end{cases}$$

are soluție deoarece:

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

(Condiția ca determinantul matricei coeficienților să fie 0.)

Rezolvarea sistemelor liniare – Istoric

Leibniz era convins că o notație matematică bună este cheia progresului și a experimentat mai mult de 50 de moduri diferite de a scrie coeficienții unui sistem de ecuații.
Leibniz folosea termenul de „rezultant” în loc de „determinant” și a demonstrat regula lui Cramer pentru „rezultanți”.

El știa că orice determinant poate fi dezvoltat în raport cu o coloană – operația se numește astăzi *dezvoltarea Laplace*.

Rezolvarea sistemelor liniare – Istoric

- 1750, Cramer prezintă o formulă bazată pe determinanți pentru rezolvarea unui sistem de ecuații liniare – *regula lui Cramer* – în *„Introduction in the analysis of algebraic curves"*.
  (dă o regulă generală pentru sisteme $n \times n$:
  „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, Bézout; 1771, Vandermonde; 1772, Laplace – reguli de calcul al determinanților;

Rezolvarea sistemelor liniare – Istoric

- 1773, Lagrange – prima utilizare implicită a matricelor în legătură cu formele biliniare ce apar la optimizarea unei funcții reale de 2 sau mai multe variabile (dorea să caracterizeze punctele de maxim și minim ale funcțiilor de mai multe variabile).

Rezolvarea sistemelor liniare – Istoric

- 1800–1801, Gauss introduce noțiunea de „determinant” (determină proprietățile formei pătratice) – *Disquisitiones arithmeticae* (1801);
- descrie operațiile de înmulțire matricială și inversă a unei matrice în contextul tabloului coeficienților unei forme pătratice;
- Gauss dezvoltă eliminarea gaussiană în timp ce studia orbita asteroidului Pallas, de unde obține un sistem liniar cu 6 ecuații și 6 necunoscute.

Rezolvarea sistemelor liniare – Istoric

- 1800–1801, Gauss introduce noțiunea de „determinant” (determină proprietățile formei pătratice) – *Disquisitiones arithmeticae* (1801);

  Acesta descrie operațiile de înmulțire matricială și inversă a unei matrice în contextul tabloului coeficienților unei forme pătratice;

  Gauss dezvoltă eliminarea gaussiană în timp ce studia orbita asteroidului Pallas, de unde obține un sistem liniar cu 6 ecuații și 6 necunoscute.

Rezolvarea sistemelor liniare – Istoric

- 1812, Cauchy folosește termenul de „determinant” în sensul cunoscut astăzi.
- 1826, Cauchy găsește valorile proprii și deduce rezultate legate de diagonalizarea unei matrice. Introduce noțiunea de matrice asemenea și demonstrează că acestea au aceeași ecuație caracteristică.

  Demonstrează că orice matrice reală simetrică este diagonalizabilă.

Rezolvarea sistemelor liniare – Istoric

- 1850, Sylvester introduce pentru prima dată termenul de „matrice” (din latină, „uter” – un loc unde ceva se formează sau este produs, „an oblong arrangement of terms”).
- 1855, Cayley – dezvoltă algebra matricială și oferă prima definiție abstractă a unei matrice.

  Studiază transformările liniare și compunerea lor, ceea ce îl conduce la operațiile cu matrice (adunare, înmulțire, înmulțirea cu un scalar, inversa).

Rezolvarea sistemelor liniare – Istoric

- 1858, Cayley, în *Memoriu asupra teoriei matricelor*: *„Sunt multe lucruri de spus despre această teorie a matricelor și, după părerea mea, această teorie ar trebui să preceadă teoria determinanților."*
- Jordan (1870 – *Treatise on substitutions and algebraic equations*) – forma canonică Jordan;
- Frobenius (1878 – *On linear substitutions and bilinear forms*) – noțiunea de rang al unei matrici;
- 1890, Weierstrass – *On determinant theory* – definiția axiomatică a determinantului.

Rezolvarea sistemelor liniare – Istoric

- 1925, Heisenberg reinventează algebra matricială pentru mecanica cuantică;
- 1947, von Neumann & Goldstine introduc numerele de condiționare atunci când analizează erorile de rotunjire;
- 1948, Turing introduce descompunerea LU a unei matrice;
- 1958, Wilkinson dezvoltă factorizarea QR;
- ...

Inversabilitatea matricelor de forma $I_n \pm A$

Propoziția 1

Fie $A \in \mathbb{R}^{n\times n}$ pentru care există o normă matriceală naturală astfel încât $\|A\| < 1$.
Atunci există matricele $(I_n \pm A)^{-1}$ și avem estimările:

$$\frac{1}{1+\|A\|} \leq \|(I_n \pm A)^{-1}\| \leq \frac{1}{1-\|A\|}.$$

Evaluarea erorii în rezolvarea sistemelor liniare

Fie $A \in \mathbb{R}^{n \times n}, \mathbf{x} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^n$ și sistemul de ecuații liniare:

$$A\mathbf{x} = \mathbf{b}.$$

Dacă A este nesingulară ($\det A \neq 0$), atunci sistemul are soluție unică:

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Pentru erorile în datele de intrare folosim notațiile:

- $\Delta A \in \mathbb{R}^{n \times n}$ – *eroarea absolută pentru matricea A;*
- $\Delta \mathbf{b} \in \mathbb{R}^n$ – *eroarea absolută pentru vectorul* $\mathbf{b}$.

Evaluarea erorii

În realitate se rezolvă sistemul:

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

soluția fiind:

$$\tilde{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}.$$

În mod natural se ridică următoarele probleme:

1. Dacă A este matrice nesingulară, pentru ce ΔA este $A + \Delta A$ nesingulară?
2. Presupunând că A și $A + \Delta A$ sunt nesingulare, care sunt relațiile între

$$\frac{\|\Delta A\|}{\|A\|},\ \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \quad \text{și} \quad \frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\,?$$

Condiție de nesingularitate

Presupunem A nesingulară.

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

Rezultă:

$A + \Delta A$ este nesingulară $\iff$ $I_n + A^{-1}\Delta A$ este nesingulară.

Propoziția 2

Fie A nesingulară și $\|\Delta A\| < \dfrac{1}{\|A^{-1}\|}$. Atunci $I_n + A^{-1}\Delta A$ este nesingulară și avem estimarea:

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

Demonstrație. Avem:

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

Din Propoziția 1 avem că $\exists\, (I_n + A^{-1}\Delta A)^{-1}$ şi are loc

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

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

Evaluarea erorii în soluție

$$(A+\Delta A)(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{b}+\Delta\mathbf{b} \Rightarrow (A+\Delta A)\Delta\mathbf{x}+A\mathbf{x}+\Delta A\,(\mathbf{x}) = \mathbf{b}+\Delta\mathbf{b}.$$

$$A(I_n + A^{-1}\Delta A)\Delta\mathbf{x} = \Delta\mathbf{b} - (\Delta A)\mathbf{x}$$

$$\Rightarrow \Delta\mathbf{x} = (I_n + A^{-1}\Delta A)^{-1}A^{-1}[\Delta\mathbf{b} - (\Delta A)\mathbf{x}]$$

$$\|\Delta\mathbf{x}\| \leq \|(I_n + A^{-1}\Delta A)^{-1}\| \cdot \|A^{-1}\| \cdot (\|\Delta\mathbf{b}\| + \|\Delta A\| \cdot \|\mathbf{x}\|)$$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|\Delta A\|}\left(\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{x}\|} + \|\Delta A\|\right)$$

Din $A\mathbf{x} = \mathbf{b}$ obținem

$$\|\mathbf{b}\| = \|A\mathbf{x}\| \leq \|A\|\,\|\mathbf{x}\| \Rightarrow \frac{1}{\|\mathbf{x}\|} \leq \frac{\|A\|}{\|\mathbf{b}\|}. \quad (*)$$

Ținând seama de $(*)$, deducem:

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\| \cdot \|A\|}{1 - \|A^{-1}\| \cdot \|\Delta A\|} \left(\frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|\Delta A\|}{\|A\|} \right)$$

- $k(A) = \|A^{-1}\| \cdot \|A\|$ - se numeşte *numărul de condiţionare* al matricei A.

Propoziţia 3

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

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

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

Numărul de condiţionare verifică

$$k(A) \geq 1, \ \forall A,$$

$k(A)$ - depinde de norma matricială naturală utilizată.

O matrice A pentru care numărul de condiționare este mare se numește *matrice prost condiționată* ($k(A)$ „mare”).

Pentru sistemul

$$Ax = b,$$

dacă avem $k(A)$ prea mare, atunci

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|}$$

poate fi mare chiar dacă erorile relative în date

$$\frac{\|\Delta A\|}{\|A\|}, \qquad \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|}$$

sunt mici.

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

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

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

Matricea simetrică A are valorile proprii reale $\lambda_1, \lambda_2, \ldots, \lambda_n$..

- A^2 are valorile proprii $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$.
- A^{-1} are valorile proprii $\dfrac{1}{\lambda_1}, \dfrac{1}{\lambda_2}, \ldots, \dfrac{1}{\lambda_n}$.

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

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

Numărul de condiţionare spectral:

$$k_2(A) = \|A\|_2 \, \|A^{-1}\|_2 = \frac{|\lambda_n|}{|\lambda_1|}.$$

Dacă A este matrice ortogonală, atunci $k_2(A) = 1$.

Cum $A^T A = AA^T = I_n$, avem $A^{-1} = A^T$.

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

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

Exemplu de matrice aproape singulară, dar cu număr de condiționare mic:

$$A = \mathrm{diag}[1, 0.1, 0.1, \ldots, 0.1] \in \mathbb{R}^{100\times 100}$$

$$\det(A) = 1 \cdot (0.1)^{99} = 10^{-99}.$$

Pe de altă parte

$$A^{-1} = \mathrm{diag}[1, 10, 10, \ldots, 10].$$

iar

$$\|A\|_2 = 1, \qquad \|A^{-1}\|_2 = 10$$

$$k_2(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = 10.$$

Matrice foarte prost condiționată cu determinant nenul ($\det A = 1$):

$$A = \begin{pmatrix} 1 & 2 & 0 & \dots & 0 \\ 0 & 1 & 2 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 2 \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

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

$$\|A\|_\infty = \|A\|_1 = 3,\ \|A^{-1}\|_\infty = \|A\|_1 = 1+2+2^2+\cdots+2^{n-1} = 2^n-1.$$

$$k(A) = \|A\|_\infty \|A^{-1}\|_\infty = \|A\|_1 \|A^{-1}\|_1 = 3(2^n - 1).$$

Pentru $n = 100$:

$$k(A) = 3(2^{100} - 1).$$

$$\det(A) = 1.$$

Exemplu – Sistem sensibil la perturbări

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

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

$$\begin{cases} 400x - 201y = 200, \\ -800x + 401y = -200 \end{cases} \qquad \begin{cases} 401x - 201y = 200, \\ -800x + 401y = -200 \end{cases}$$

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

$$k_2(A) = 2503 \qquad\qquad k_2(A) = 1002000$$

Exemplu – Sistem prost condiționat

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

$$x = 2, \qquad y = -2. \text{ iar } \quad \overline{x} = 0.9911,\ \overline{y} = -0.4870.$$

$$r = b - Az = \begin{pmatrix} 0.8642 \\ 0.1440 \end{pmatrix} - \begin{pmatrix} 1.2969 & 0.8648 \\ 0.2161 & 0.1441 \end{pmatrix} \begin{pmatrix} 0.9911 \\ 0.4870 \end{pmatrix} = \begin{pmatrix} 10^{-8} \\ 10^{-8} \end{pmatrix}.$$

Matricea Hilbert

Definiție. Matricea Hilbert $H_n = (h_{ij})_{i,j=1}^{n}$ este definită prin:

$$h_{ij} = \frac{1}{i+j-1} = \int_0^1 x^{i+j-2}\, dx, \qquad i, j = 1, \ldots, n.$$

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

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

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