Universitatea din Craiova

Facultatea de Ştiinţe

Specializarea: Matematică

Cicul I - Licență - ANUL III

# Algoritmică şi Simulare Numerică in C++

*Autor:*

Prof. Dr. Sorin Micu

2026

# Cursul 1

# Elemente de teorie matricială (1)

## Mulţimi de numere

$\mathbb{N}$ Numere naturale
$\mathbb{Z}$ Numere întegi
$\mathbb{R}$ Numere reale
$\mathbb{C}$ Numere complexe

## Vectori

Un vector linie $x$ şi un vector coloană $y$ de dimensiune $n$:

$$x = (x_1, x_2, \ldots, x_n) \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Mulţimea vectorilor reali (complecşi) coloană de dimensiune $n$ se notează $\mathbb{R}^n$ ($\mathbb{C}^n$). Componenta $i$ a lui $y$ se notează, ca mai sus, $y_i$, şi este un număr real (complex).

Operaţii cu vectori:

1. $y = (y_i)_{1\le i\le n} \in \mathbb{C}^n,\ z = (z_i)_{1\le i\le n} \in \mathbb{C}^n$: $y + z = (y_i + z_i)_{1\le i\le n} \in \mathbb{C}^n$.
2. $y = (y_i)_{1\le i\le n} \in \mathbb{C}^n,\ \alpha \in \mathbb{C}$: $\alpha y = (\alpha y_i)_{1\le i\le n} \in \mathbb{C}^n$.

$(\mathbb{C}^n, +, \cdot)$ este un spaţiu vectorial.

- O mulţime de vectori $\{v^1, v^2, ..., v^m\} \subset \mathbb{C}^n$ este *liniar independentă* dacă

$$\sum_{i=1}^{m} \alpha_i v^i = 0 \Rightarrow \alpha_i = 0,\ 1 \le i \le m.$$

- O mulţime de vectori $\{v^1, v^2, ..., v^m\} \subset \mathbb{C}^n$ este *sistem de generatori în* $\mathbb{C}$ dacă, pentru fiecare $v \in \mathbb{C}^n$, există scalari $(\alpha_i)_{1\le i\le m} \subset \mathbb{C}$ astfel încât

$$v = \sum_{i=1}^{m} \alpha_i v^i.$$

- O mulţime de vectori $\{v^1, v^2, ..., v^m\} \subset \mathbb{C}^n$ este *bază* dacă este liniar independentă şi sistem de generatori. Numărul de vectori dintr-o bază a lui $\mathbb{C}^n$ se numeşte *dimensiunea spaţiului vectorial* $\mathbb{C}^n$.

**Teorema 1.1.**

1. *O mulţime de vectori $\{v^1, v^2, ..., v^m\} \subset \mathbb{C}^n$ este bază a lui $\mathbb{C}^n$ dacă, pentru fiecare $v \in \mathbb{C}^n$, există scalarii $(\alpha_i)_{1\le i\le m} \subset \mathbb{C}$ unici astfel încât $v = \sum_{i=1}^m \alpha_i v^i$.*
2. *O mulţime de $n$ vectori liniar independenţi din $\mathbb{C}^n$ formează o bază. Dimensiunea lui $\mathbb{C}^n$ este egală cu $n$.*

- Mulţimea următorilor vectori formează o bază a lui $\mathbb{C}^n$ (numită *baza canonică*):

$$e^1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e^2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad e^n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

## Matrici

O matrice $m \times n$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} = (a_{ij})_{1\le i\le m;\, 1\le j\le n}.$$

Primul indice este de linie iar al doilea de coloană. Mulţimea matricilor reale (complexe) de dimensiune $m \times n$ se notează $\mathcal{M}_{m\times n}(\mathbb{R})$ ($\mathcal{M}_{m\times n}(\mathbb{C})$). Elementul $(i, j)$ al lui $A$ se notează, ca mai sus, $a_{ij}$, şi este un număr real (complex). Dacă $m = n$ (numărul de linii este egal cu cel de coloane), matricea se numeşte pătratică. Mulţimea matricilor reale (complexe) pătratice de dimensiune $n$ se notează $\mathcal{M}_n(\mathbb{R})$ ($\mathcal{M}_n(\mathbb{C})$).

- Operaţii cu matrici:

1. $A = (a_{ij})_{1\le i\le m;\, 1\le j\le n} \in \mathcal{M}_{m\times n}(\mathbb{C}),\ B = (b_{ij})_{1\le i\le m;\, 1\le j\le n} \in \mathcal{M}_{m\times n}(\mathbb{C}),$

$$A + B = (a_{ij} + b_{ij})_{1\le i\le m;\, 1\le j\le n}.$$

2. $A = (a_{ij})_{1\le i\le m;\, 1\le j\le n} \in \mathcal{M}_{m\times n}(\mathbb{C}),\ \alpha \in \mathbb{C},$

$$\alpha A = (\alpha a_{ij})_{1\le i\le m;\, 1\le j\le n}.$$

3. $A = (a_{ij})_{1\le i\le m;\, 1\le j\le n} \in \mathcal{M}_{m\times n}(\mathbb{C}),\ B = (b_{ij})_{1\le i\le n;\, 1\le j\le p} \in \mathcal{M}_{n\times p}(\mathbb{C}),$

$$AB = \left(\sum_{k=1}^n a_{ik} b_{kj}\right)_{1\le i\le m;\, 1\le j\le p} \in \mathcal{M}_{m\times p}(\mathbb{C}).$$

Avem că $(\mathcal{M}_{m\times n}(\mathbb{C}), +, \cdot)$ este un spaţiu vectorial.

- Matricea identitate şi matrice diagonală în $\mathcal{M}_n(\mathbb{C})$:

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix} = \text{diag}[d_1, d_2, \dots, d_n], \ \ d_i \in \mathbb{C},\ 1 \leq i \leq n.$$

- Matrici triunghiulare în $\mathcal{M}_n(\mathbb{C})$:

$$A = \begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-11} & a_{n-12} & a_{n-13} & \dots & 0 \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}, \quad \text{matrice triunghiulară inferior: } a_{ij} = 0 \text{ dacă } j > i;$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}, \quad \text{matrice triunghiulară superior: } a_{ij} = 0 \text{ dacă } i > j;$$

- Matrice conjugată (adjunctă). Dată $A = (a_{ij})_{1 \leq i \leq m;\, 1 \leq j \leq n} \in \mathcal{M}_{m \times n}(\mathbb{C})$, matricea sa conjugată se defineşte prin

$$A^* = (\overline{a}_{ji})_{1 \leq j \leq n;\, 1 \leq i \leq m} \in \mathcal{M}_{n \times m}(\mathbb{C}).$$

  Dacă notăm $A^t$ transpusa lui $A$ (matricea în care liniile lui $A$ devin coloane) atunci $A^*$ este matricea ale cărei elemente sunt conjugatele complexe ale elementelor lui $A^t$. Dacă $A$ are doar elemente reale, atunci $A^* = A^t$.

  Avem următoarea proprietate: $(AB)^* = B^* A^*$.

- Determinantul unei matrici pătratice $A = (a_{ij})_{1 \leq i \leq n} \in \mathcal{M}_n(\mathbb{C})$ este dat de

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}.$$

  Avem următoarea proprietate: $\det(AB) = \det(A)\det(B)$.

- Matrice normală, hermitică (simetrică), unitară (ortogonală), definită pozitiv. Dată o matrice pătratică $A \in \mathcal{M}_n(\mathbb{C})$:

  - $A$ este normală dacă $AA^* = A^* A$;
  - $A$ este hermitică dacă $A = A^*$;

    Dacă $A \in \mathcal{M}_n(\mathbb{R})$ şi $A = A^t$ atunci matricea $A$ se numeşte simetrică.
  - $A$ este unitară dacă $AA^* = A^* A = I$;

    Dacă $A \in \mathcal{M}_n(\mathbb{R})$ şi $AA^t = A^t A = I$ atunci matricea $A$ se numeşte ortogonală.
  - O matrice hermitică $A$ este definită pozitiv dacă

$$v^* A v > 0 \quad \forall v \in \mathbb{C}^n \setminus \{0\}.$$

- O matrice pătratică $A \in \mathcal{M}_n(\mathbb{C})$ este inversabilă dacă există $A^{-1} \in \mathcal{M}_n(\mathbb{C})$ astfel ca

$$AA^{-1} = A^{-1}A = I.$$

**Teorema 1.2.** *O matrice $A \in \mathcal{M}_n(\mathbb{C})$ este inversabilă dacă şi numai dacă $det(A) \neq 0$.*

- Matrice diagonalizabilă. O matrice pătratică $A \in \mathcal{M}_n(\mathbb{C})$ este diagonalizabilă dacă există o matrice inversabilă $P \in \mathcal{M}_n(\mathbb{C})$ şi o matrice diagonală $D \in \mathcal{M}_n(\mathbb{C})$ astfel ca

$$P^{-1}AP = D.$$

**Teorema 1.3.**

1. *Dacă $A$ este o matrice normală atunci există o matrice unitară $U$ astfel încât $U^*AU$ este diagonală.*
2. *Dacă $A$ este o matrice pătratică oarecare atunci există o matrice unitară $U$ astfel încât $U^*AU$ este triunghiulară superior.*

## Produsul scalar canonic.

Aplicaţia $(\,\cdot\,,\,\cdot\,) : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ definită prin:

$$(u, v) = v^*u = \sum_{i=1}^{n} u_i \overline{v}_i$$

se numeşte produsul scalar canonic în $\mathbb{C}^n$. Norma euclidiană este definită de

$$\|v\|_2 = \sqrt{(v, v)} = \sqrt{\sum_{i=1}^{n} |v_i|^2} \quad \forall v \in \mathbb{C}^n.$$

**Teorema 1.4** (Proprietăţile produsului scalar)**.**

1. *(Pozitivitate) $(v, v) \geq 0$ pentru orice $v \in \mathbb{C}^n$ şi $(v, v) = 0$ doar dacă $v = 0$;*
2. *(Liniaritate) $(\alpha u + \beta v, w) = \alpha(u, w) + \beta(v, w)$ pentru orice $u$, $v$, $w \in \mathbb{C}^n$ şi $\alpha$, $\beta \in \mathbb{C}$;*
3. *(Antisimetrie) $(u, v) = \overline{(v, u)}$ pentru orice $u$, $v \in \mathbb{C}^n$ .*

- Doi vectori $u$ şi $v$ din $\mathbb{C}^n$ se numesc ortogonali dacă $(u, v) = 0$. Un vector $u$ din $\mathbb{C}^n$ se numeşte unitar dacă $\|u\|_2 = 1$.

## Valori şi vectori proprii.

Fie o matrice pătratică $A \in \mathcal{M}_n(\mathbb{C})$. $\lambda \in \mathbb{C}$ este valoare proprie a lui $A$ dacă există $v \in \mathbb{C}^n$, $v \neq 0$, astfel încât

$$Av = \lambda v.$$

Un vectorul $v$ cu proprietatea de mai sus se numeşte vector propriu al lui $A$ corespunzător valorii proprii $\lambda$.

**Teorema 1.5.** *$\lambda$ este valoare proprie a lui $A$ dacă şi numai dacă $det(A-\lambda I)=0$ ($\lambda$ este rădăcină a polinomului caracteristic).*

- Spectrul matricii pătratice $A$ este

$$\sigma(A)=\{\lambda\,:\,\lambda \text{ este valoare proprie a lui } A\}.$$

- Raza spectrală a matricii pătratice $A$ este

$$\rho(A)=\sup\{|\lambda|\,:\,\lambda \text{ este valoare proprie a lui } A\}.$$

**Teorema 1.6.** *Fiind dată o matrice pătratică $A\in\mathcal{M}_n(\mathbb{C})$ şi o matrice inversabilă $P\in\mathcal{M}_n(\mathbb{C})$, avem că*

$$\sigma(A)=\sigma(P^{-1}AP).$$

- O mulţime $\{v^1,v^2,...,v^m\}$ a lui $\mathbb{C}^n$ este *ortonormată* dacă:

$$(v^i,v^j)=\delta_{ij}\quad 1\le i,j\le m.$$

**Teorema 1.7.** *Fie $\{v^1,v^2,...,v^n\}$ o mulţime ortonormată a lui $\mathbb{C}^n$.*

1. *$\{v^1,v^2,...,v^n\}$ este bază a lui $\mathbb{C}^n$;*
2. *Orice vector $u\in\mathbb{C}^n$ se scrie:*
$$u=\sum_{i=1}^{n}(u,v^i)v^i.$$
*$(u,v^i)$ sunt coeficienţii lui $u$ în baza ortonormată $\{v^1,v^2,...,v^n\}$ a lui $\mathbb{C}^n$.*

**Teorema 1.8.** *Se poate forma o bază ortonormată a lui $\mathbb{C}^n$ din vectori proprii ai oricărei matrici normale.*

*Proof.* Fie $A\in\mathcal{M}_n(\mathbb{C})$ o matrice normală. Conform Teoremei 1.3, ea este diagonalizabilă deci există o matrice unitară $U\in\mathcal{M}_n(\mathbb{C})$ şi o matrice diagonală $D\in\mathcal{M}_n(\mathbb{C})$ astfel ca

$$U^*AU=D=\text{diag}[\lambda_1,\lambda_2,\ldots,\lambda_n].$$

Dacă notăm $v^i$ coloana $i$ a matricii $U$ atunci din faptul că $U$ este unitară rezultă că

$$(v^i,v^j)=(v^j)^*v^i=(U^*U)_{ji}=I_{ji}=\delta_{ij},$$

deci mulţimea $\{v^1,v^2,...,v^n\}$ a lui $\mathbb{C}^n$ este bază ortonormată a lui $\mathbb{C}^n$.

Din relaţia de diagonalizare deducem că

$$(U^*AU)_{ii}=D_{ii}\Leftrightarrow(v^i)^*Av^i=\lambda_i\Leftrightarrow Av^i=\lambda_iv^i\quad 1\le i\le n,$$

deci $v^i$ este vector propriu al lui $A$ corespunzător valorii proprii $\lambda_i$. □

**Teorema 1.9** (Teorema Courant-Fischer)**.** *Fie $A \in \mathcal{M}_n(\mathbb{C})$ o matrice hermitică şi $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ valorile sale proprii. Pentru fiecare $v \in \mathbb{C}^n \setminus \{0\}$, fie $R_A(v) = \frac{v^*Av}{v^*v}$. Avem că:*

1. $\lambda_n = \max_{v \in \mathbb{C}^n \setminus \{0\}} R_A(v)$;
2. $\lambda_1 = \min_{v \in \mathbb{C}^n \setminus \{0\}} R_A(v)$.

*Proof.* Conform Teoremei 1.8, matricea $A$ fiind normală, poate fi diagonalizată prin intermediul unei matrici unitare $U$, ale cărei coloane sunt vectori proprii ai lui $A$ şi formează o bază ortonormată a lui $\mathbb{C}^n$. Fie $p^1, p^2, \ldots, p^n$ aceste coloane, care sunt şi vectori proprii corespunzători valorilor proprii $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ ale lui $A$. Pentru fiecare vector $v \in \mathbb{C}^n$ există scalarii $v_1, v_2, \ldots, v_n$ astfel ca $v = \sum_{i=1}^n v_i p^i$. Avem că

$$R_A(v) = \frac{v^*Av}{v^*v} = \frac{(Av, v)}{(v, v)} = \frac{\left(A\left(\sum_{i=1}^n v_i p^i\right), \sum_{i=1}^n v_i p^i\right)}{\left(\sum_{i=1}^n v_i p^i, \sum_{i=1}^n v_i p^i\right)} = \frac{\left(\sum_{i=1}^n \lambda_i v_i p^i, \sum_{i=1}^n v_i p^i\right)}{\left(\sum_{i=1}^n v_i p^i, \sum_{i=1}^n v_i p^i\right)} =$$

$$= \frac{\sum_{i=1}^n \sum_{j=1}^n \left(\lambda_i v_i p^i, v_j p^j\right)}{\sum_{i=1}^n \sum_{j=1}^n \left(v_i p^i, v_j p^j\right)} = \frac{\sum_{i=1}^n \sum_{j=1}^n \lambda_i v_i \overline{v_j} \left(p^i, p^j\right)}{\sum_{i=1}^n \sum_{j=1}^n v_i \overline{v_j} \left(p^i, p^j\right)} = \frac{\sum_{i=1}^n \lambda_i v_i \overline{v_i}}{\sum_{i=1}^n v_i \overline{v_i}} = \frac{\sum_{i=1}^n \lambda_i |v_i|^2}{\sum_{i=1}^n |v_i|^2}.$$

Ţinând cont de ordonarea valorilor proprii, avem că

$$R_A(v) = \frac{\sum_{i=1}^n \lambda_i |v_i|^2}{\sum_{i=1}^n |v_i|^2} \leq \frac{\sum_{i=1}^n \lambda_n |v_i|^2}{\sum_{i=1}^n |v_i|^2} = \lambda_n,$$

$$R_A(v) = \frac{\sum_{i=1}^n \lambda_i |v_i|^2}{\sum_{i=1}^n |v_i|^2} \geq \frac{\sum_{i=1}^n \lambda_1 |v_i|^2}{\sum_{i=1}^n |v_i|^2} = \lambda_1.$$

Cum $R_A(p^1) = \lambda_1$ şi $R_A(p^n) = \lambda_n$ obţinem cele două relaţii din teoremă. □

## Probleme

1. Găsiţi o bază a spaţiului $\mathcal{M}_{m \times n}(\mathbb{C})$. Care este dimensiunea acestui spaţiu?
2. Pentru orice matrice $A \in \mathcal{M}_n(\mathbb{C})$ şi orice doi vectori $v, w \in \mathbb{C}^n$ avem că: $w^*Av = (A^*w)^*v$, $(Av, w) = (v, A^*w)$, $(v, \alpha w) = \overline{\alpha}(v, w)$.
3. Dacă $A \in \mathcal{M}_n(\mathbb{C})$ este hermitică atunci $v^*Av \in \mathbb{R}$ pentru orice $v \in \mathbb{C}^n$.
4. Pentru orice $A \in \mathcal{M}_n(\mathbb{C})$ avem că $A^*A$ este hermitică.
5. Pentru orice $A \in \mathcal{M}_n(\mathbb{C})$ şi orice $v \in \mathbb{C}^n$ avem că $v^*A^*Av \geq 0$. Este $A^*A$ definită pozitiv?
6. Valorile proprii ale unei matrici hermitice sunt reale. Vectorii proprii ai unei matrici hermitice corespunzători la valori proprii diferite sunt ortogonali.
7. Dacă $U \in \mathcal{M}_n(\mathbb{C})$ este o matrice unitară atunci $\|Uv\|_2 = \|v\|_2$ pentru orice $v \in \mathbb{C}^n$.
8. Valorile proprii ale unei matrici unitare au modulul egal cu 1.
9. O matrice diagonală este definită pozitiv dacă şi numai dacă are elementele de pe diagonala principală reale şi pozitive.
10. O matrice hermitică este definită pozitiv dacă şi numai dacă are toate valorile proprii reale şi pozitive.

# Cursul 2

# Elemente de teorie matricială (2)

## 2.1 Norme vectoriale în $\mathbb{C}^n$.

O aplicaţie $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ este o **normă vectorială în** $\mathbb{C}^n$ dacă are următoarele proprietăţi:

1. (Pozitivitate) $\|v\| \geq 0$ pentru orice $v \in \mathbb{C}^n$ şi $\|v\| = 0$ doar dacă $v = 0$;
2. (Omogeneitate) $\|\alpha u\| = |\alpha|\|u\|$ pentru orice $u \in \mathbb{C}^n$ şi $\alpha \in \mathbb{C}$;
3. (Inegalitatea triunghiului) $\|u + v\| \leq \|u\| + \|v\|$ pentru orice $u$, $v \in \mathbb{C}^n$.

- Exemple:

1. Norma euclidiană (doi):

$$\|v\|_2 = \sqrt{(v, v)} = \sqrt{\sum_{i=1}^{n} |v_i|^2} \quad \forall v \in \mathbb{C}^n.$$

2. Norma infinit:

$$\|v\|_\infty = \max_{1 \leq i \leq n} |v_i| \quad \forall v \in \mathbb{C}^n.$$

3. Norma unu:

$$\|v\|_1 = \sum_{i=1}^{n} |v_i| \quad \forall v \in \mathbb{C}^n.$$

## 2.2 Norme matriciale în $\mathcal{M}_n(\mathbb{C})$:

O aplicaţie $\|\cdot\| : \mathcal{M}_n(\mathbb{C}) \to \mathbb{R}$ este o **normă matricială în** $\mathcal{M}_n(\mathbb{C})$ dacă are următoarele proprietăţi:

1. (Pozitivitate) $\|A\| \geq 0$ pentru orice $A \in \mathcal{M}_n(\mathbb{C})$ şi $\|A\| = 0$ doar dacă $A = 0$;
2. (Omogeneitate) $\|\alpha A\| = |\alpha|\|A\|$ pentru orice $A \in \mathcal{M}_n(\mathbb{C})$ şi $\alpha \in \mathbb{C}$;
3. (Inegalitatea triunghiului) $\|A + B\| \leq \|A\| + \|B\|$ pentru orice $A$, $B \in \mathcal{M}_n(\mathbb{C})$;
4. (Inegalitatea produsului) $\|AB\| \leq \|A\|\|B\|$ pentru orice $A$, $B \in \mathcal{M}_n(\mathbb{C})$.

**Teorema 2.1.** *Fie $\|\cdot\|:\mathbb{C}^n\to\mathbb{R}$ o normă vectorială în $\mathbb{C}^n$. Aplicaţia $\|\cdot\|:\mathcal{M}_n(\mathbb{C})\to\mathbb{R}$ definită prin*

$$\|A\| = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|}{\|v\|} \tag{2.1}$$

*este o normă matricială numită* **norma matricială subordonată normei vectoriale** $\|\cdot\|$.

*Proof.* Verificăm cele patru proprietăţi care definesc norma matricială:

1. $\|A\| = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|}{\|v\|} \geq 0$

   Dacă $\sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|}{\|v\|} = 0$ rezultă că $Av = 0$ pentru orice $v\in\mathbb{C}^n$, deci $A = 0$.

2. $\|\alpha A\| = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|\alpha Av\|}{\|v\|} = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{|\alpha|\|Av\|}{\|v\|} = |\alpha| \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|}{\|v\|} = |\alpha|\|A\|$.

3. $\|A+B\| = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|(A+B)v\|}{\|v\|} \leq \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|+\|Bv\|}{\|v\|}$
   $\leq \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|}{\|v\|} + \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Bv\|}{\|v\|} = \|A\| + \|B\|$.

4. $\|AB\| = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|(AB)v\|}{\|v\|} = \sup_{v\in\mathbb{C}^n,\, Bv\neq 0} \frac{\|A(Bv)\|}{\|Bv\|}\frac{\|Bv\|}{\|v\|}$
   $\leq \sup_{v\in\mathbb{C}^n,\, Bv\neq 0} \frac{\|A(Bv)\|}{\|Bv\|} \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Bv\|}{\|v\|} = \|B\| \sup_{v\in\mathbb{C}^n,\, Bv\neq 0} \frac{\|A(Bv)\|}{\|Bv\|}$
   $\leq \|B\| \sup_{w\in\mathbb{C}^n,\, w\neq 0} \frac{\|Aw\|}{\|w\|} = \|A\|\|B\|$.

□

**Observaţia 2.1.** *Avem următoarele egalităţi:*

$$\sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|}{\|v\|} = \sup_{v\in\mathbb{C}^n,\ \|v\|\leq 1,\ v\neq 0} \frac{\|Av\|}{\|v\|} = \sup_{v\in\mathbb{C}^n,\ \|v\|\leq 1} \|Av\| = \sup_{v\in\mathbb{C}^n,\ \|v\|=1} \|Av\|.$$

**Observaţia 2.2.** *Avem că $\sup_{v\in\mathbb{C}^n\setminus\{0\}} \frac{\|Av\|}{\|v\|} = \beta$ dacă şi numai dacă numărul real $\beta$ are următoarele proprietăţi:*

1. *$\|Av\| \leq \beta\|v\| \quad v\in\mathbb{C}^n$;*
2. *Există $v\in\mathbb{C}^n\setminus\{0\}$ astfel încât $\|Av\| = \beta\|v\|$.*

*Într-adevăr, dacă se verifică 1.-2., atunci conform lui 1. $\beta$ este un majorant pentru mulţimea $\left\{\frac{\|Av\|}{\|v\|} : v\in\mathbb{C}^n\setminus\{0\}\right\}$ iar conform lui 2. este cel mai mic majorant.*

*Reciproc, dacă $\sup_{v\in\mathbb{C}^n\setminus\{0\}} \frac{\|Av\|}{\|v\|} = \beta$ atunci 1. se verifică deoarece $\beta$ este un majorant al mulţimii $\left\{\frac{\|Av\|}{\|v\|} : v\in\mathbb{C}^n\setminus\{0\}\right\}$ iar pentru $v = 0$ inegalitatea se verifică în mod trivial. Ramâne de demonstrat 2. Aplicaţia*

$$\|\cdot\|_A : \mathbb{C}^n\to\mathbb{R}, \quad v\to\|v\|_A := \|Av\|$$

*este continuă iar mulţimea $S(0,1) = \{v\in\mathbb{C}^n : \|v\| = 1\}$ este compactă, prin urmare există $u\in S(0,1)$ astfel ca*

$$\|v\|_A = \sup_{w\in\mathbb{C}^n,\, \|w\|=1} \|w\|_A = \sup_{w\in\mathbb{C}^n,\, \|w\|=1} \|Aw\| = \beta = \beta\|v\|.$$

- Exemple de norme matriciale:

  1. Norma doi:
  $$\|A\|_2 = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|_2}{\|v\|_2}.$$

  2. Norma infinit:
  $$\|A\|_\infty = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|_\infty}{\|v\|_\infty}.$$

  Se poate calcula după formula:
  $$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|\,.$$

  Pentru a arăta această relaţie fie $\beta = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$ şi folosim Observaţia 2.2:
  - Pentru orice $v \in \mathbb{C}^n$, avem că
  $$\|Av\|_\infty = \max_{1\leq i\leq n} \left|\sum_{j=1}^{n} a_{ij}v_j\right| \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|\,\|v\|_\infty = \beta\|v\|_\infty.$$
  - Fie $i_0$ cu proprietatea că $\beta = \sum_{j=1}^{n} |a_{i_0 j}|$ şi fie $u = (u_j)_{1\leq j\leq n} \in \mathbb{C}^n$ definit de
  $$u_j = \begin{cases} \dfrac{\overline{a_{i_0 j}}}{|a_{i_0 j}|} & \text{dacă } a_{i_0 j} \neq 0 \\ \\ 1 & \text{dacă } a_{i_0 j} \neq 0 \end{cases}$$

  Avem că:
  $$\|Au\|_\infty = \max_{1\leq i\leq n} \left|\sum_{j=1}^{n} a_{ij}u_j\right| \geq \left|\sum_{j=1}^{n} a_{i_0 j}u_j\right| = \left|\sum_{j=1}^{n} |a_{i_0 j}|\right| = \beta = \beta\|u\|_\infty.$$

  3. Norma unu:
  $$\|A\|_1 = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|_1}{\|v\|_1}.$$

  Se poate calcula după formula:
  $$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|\,.$$

  4. Norma Frobenius:
  $$\|A\|_E = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}}.$$

  Este o normă matricială care nu este subordonată nici unei norme vectoriale dacă $n \geq 2$. (Calculaţi $\|I\|_E$).

**Teorema 2.2.** *Norma matricială $\|\cdot\|_2$ are următoarele proprietăţi:*

1. $\|A\|_2 = \sqrt{\rho(AA^*)} = \sqrt{\rho(A^*A)} = \|A^*\|_2$;
2. *Dacă $U$ este o matrice unitară ($UU^* = I$) atunci $\|A\|_2 = \|AU\|_2 = \|UA\|_2 = \|U^*AU\|_2$;*
3. *Dacă $A$ este normală ($AA^* = A^*A$), atunci $\|A\|_2 = \rho(A)$.*

*Proof.* 1. Avem că

$$\|A\|_2 = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\|Av\|_2}{\|v\|_2} = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \frac{\sqrt{(Av, Av)}}{\sqrt{(v, v)}} = \sup_{v\in\mathbb{C}^n,\, v\neq 0} \sqrt{\frac{v^*A^*Av}{v^*v}} = \sqrt{\sup_{v\in\mathbb{C}^n,\, v\neq 0} R_{A^*A}(v)}.$$

Cum $A^*A$ este hermitică, Teorema **??** implică faptul că

$$\sup_{v\in\mathbb{C}^n,\, v\neq 0} R_{A^*A}(v) = \lambda_n,$$

unde $\lambda_n$ este cea mai mare valoare proprie a lui $A^*A$. Cum, în plus, toate valorile proprii ale lui $A^*A$ sunt mai mari sau egale cu zero, rezultă că

$$\|A\|_2 = \sqrt{\sup_{v\in\mathbb{C}^n,\, v\neq 0} R_{A^*A}(v)} = \sqrt{\rho(A^*A)}.$$

Pornind de la $\|A^*\|_2$ se arată în acelaşi mod că

$$\|A^*\|_2 = \sqrt{\sup_{v\in\mathbb{C}^n,\, v\neq 0} R_{AA^*}(v)} = \sqrt{\rho(AA^*)}.$$

Rămâne de arătat că $\rho(A^*A) = \rho(AA^*)$. Fie $\lambda \in \sigma(A^*A)$ astfel încât $\rho(A^*A) = \lambda$. Există $v \neq 0$ astfel încât $A^*Av = \lambda v$ şi deci $AA^*(Av) = \lambda Av$.

- Dacă $\lambda \neq 0$ atunci $Av \neq 0$ (altfel $\lambda v = A^*Av = 0$, deci $\lambda = 0$). Prin urmare $\lambda$ este valoare proprie a matricii $AA^*$ iar $Av$ este un vector propriu corespunzător. Rezultă că

  $$\rho(AA^*) \geq \lambda = \rho(A^*A).$$

  Fie acum $\mu \in \sigma(AA^*)$ astfel încât $\rho(AA^*) = \mu$. Există $w \neq 0$ astfel încât $AA^*w = \mu w$ şi deci $A^*A(A^*w) = \mu A^*w$. Cum $\mu \geq \lambda > 0$, avem că $A^*w \neq 0$ (altfel $\mu w = AA^*w = 0$, deci $\mu = 0$) şi prin urmare $\mu$ este valoare proprie a matricii $A^*A$ iar $A^*w$ este un vector propriu corespunzător. Rezultă că

  $$\rho(A^*A) \geq \mu = \rho(AA^*).$$

  În concluzie, $\rho(A^*A) = \rho(AA^*)$.
- Dacă $\lambda = 0$ atunci fie $\mu \in \sigma(AA^*)$ astfel încât $\rho(AA^*) = \mu$. Există $w \neq 0$ astfel încât $AA^*w = \mu w$ şi deci $A^*A(A^*w) = \mu A^*w$. Dacă $A^*w \neq 0$, atunci $\mu$ este valoare proprie a matricii $A^*A$ ceea ce contrazice faptul că $\rho(A^*A) = 0$. Prin urmare $A^*w = 0$ şi $\mu w = AA^*w = 0$, deci $\mu = 0$. În concluzie, în acest caz avem $\rho(A^*A) = \rho(AA^*) = 0$.

2. Avem că

$$\|AU\|_2 = \sqrt{\rho(AU(AU)^*)} = \sqrt{\rho(AUU^*A^*)} = \sqrt{\rho(AA^*)} = \|A\|_2,$$
$$\|UA\|_2 = \sqrt{\rho(UA(UA)^*)} = \sqrt{\rho(UAA^*U^*)} = \sqrt{\rho(AA^*)} = \|A\|_2,$$
$$\|U^*AU\|_2 = \sqrt{\rho(U^*AU(U^*AU)^*)} = \sqrt{\rho(U^*AUU^*A^*U)} = \sqrt{\rho(U^*AA^*U)} = \sqrt{\rho(AA^*)} = \|A\|_2.$$

3. Cum $A$ este o matrice normală, există o matrice unitară $U$ şi o matrice diagonală $D$ astfel încât $U^*AU = D$. Avem că

$$\|A\|_2 = \|U^*AU\|_2 = \|D\|_2 = \sqrt{\rho(D^*D)}.$$

Cum $\sigma(A) = \sigma(D)$ rezultă că $\sigma(DD^*) = \{|\lambda|^2 : \lambda \in \sigma(A)\}$. Prin urmare

$$\rho(D^*D) = \max\{|\lambda|^2 : \lambda \in \sigma(A)\} = (\rho(A))^2,$$

de unde rezultă concluzia $\|A\|_2 = \rho(A)$.

□

**Teorema 2.3.**

1. *Pentru orice matrice $A \in \mathcal{M}_n(\mathbb{C})$ şi orice normă matricială $\|\ \|$ avem că*

$$\rho(A) \leq \|A\|.$$

2. *Pentru orice $\varepsilon > 0$ şi orice matrice $A \in \mathcal{M}_n(\mathbb{C})$ există o normă matricială $\|\ \|$ subordonată unei norme vectoriale astfel ca*

$$\|A\| \leq \rho(A) + \varepsilon.$$

*Proof.*

1. Fie $\lambda \in \sigma(A)$ cu proprietatea că $|\lambda| = \rho(A)$. Dacă $v$ este un vector propriu al lui $A$ corespunzător valorii proprii $\lambda$, fie $w \in \mathbb{C}^n$ cu proprietatea că $vw^* \neq 0$. Avem că

$$|\lambda|\|vw^*\| = \|\lambda vw^*\| = \|Avw^*\| \leq \|A\|\|vw^*\|,$$

prin urmare $\rho(A) = |\lambda| \leq \|A\|$.

2. Pentru orice matrice $A$ există o matrice inversabilă $P$ cu proprietatea că $P^{-1}AP$ este o matrice triunghiulară superior $T$:

$$P^{-1}AP = T = \begin{pmatrix} \lambda_1 & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & \lambda_2 & t_{23} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \lambda_{n-1} & t_{n-1n} \\ 0 & 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Să observăm că elementele de pe diagonala principală a lui $T$ sunt valorile proprii ale lui $A$. Pentru $\delta > 0$, definim matricea $D_\delta = \mathrm{diag}[1, \delta, \delta^2, \dots, \delta^{n-1}]$. Avem că

$$D_\delta^{-1}P^{-1}APD_\delta = \begin{pmatrix} \lambda_1 & \delta t_{12} & \delta^2 t_{13} & \dots & \delta^{n-1} t_{1n} \\ 0 & \lambda_2 & \delta t_{23} & \dots & \delta^{n-2} t_{2n} \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \lambda_{n-1} & \delta t_{n-1n} \\ 0 & 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Dat $\varepsilon > 0$ fie $\delta > 0$ astfel ca

$$\sum_{j=i+1}^{n} \delta^{j-i}|t_{ij}| \leq \varepsilon \qquad (1 \leq i \leq n-1).$$

Definim norma matricială

$$\|B\| = \|(PD_\delta)^{-1}BPD_\delta\|_\infty \quad \forall B \in \mathcal{M}_n(\mathbb{C}),$$

care este norma matricială subordonată normei vectoriale $v \to \|v\| = \|(PD_\delta)^{-1}v\|_\infty$. În plus, avem că:

$$\|A\| = \|(PD_\delta)^{-1}BPD_\delta\|_\infty = \max_{1\le i\le n}\left(|\lambda_i| + \sum_{j=i+1}^{n}\delta^{j-i}|t_{ij}|\right)$$

$$\le \max_{1\le i\le n}|\lambda_i| + \max_{1\le i\le n}\sum_{j=i+1}^{n}\delta^{j-i}|t_{ij}| \le \rho(A) + \varepsilon.$$

□

## 2.3 Exerciţii

1. $\|A\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|$
2. $\|A\|_E^2 = Tr(A^*A)$, $\|UA\|_E = \|AU\|_E = \|U^*AU\|_E$ pentru orice $U$ unitară.
3. $A \in \mathcal{M}_n(\mathbb{C})$ are proprietatea că există o normă subordonată astfel ca $\|A\| < 1$. Atunci $A - I$ este inversabilă.
4. Fie $u \in \mathbb{C}^n$ cu $\|u\|_2 = 1$. Arătaţi că $H = I - 2uu^*$ este o matrice unitară şi hermitică.
5. Fie $n \ge 1$ şi matricea $A = (a_{ij})_{1\le i,j\le n} \in \mathcal{M}_n(\mathbb{R})$ ale cărei elemente sunt date de

$$a_{ij} = \begin{cases} -1 & \text{dacă } |i-j| = 1 \\ 2 & \text{dacă } i = j \\ 0 & \text{în celelalte cazuri.} \end{cases}$$

   Arătaţi că $A$ este definită pozitiv şi găsiţi vectorii şi valorile ei proprii.

   Indicaţie: Se va arăta că vectorii proprii sunt de forma $(sin(k\pi jh))_{1\le j\le n} \in \mathbb{R}^n$ unde $k \in \{1, 2, ..., n\}$ şi $h = \frac{1}{n+1}$.
6. Arătaţi că determinantul oricărei matrici hermitice este un număr real.
7. Date $\theta, \zeta \in [0, 2\pi)$, arătaţi că matricea $A = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta)e^{-i\zeta} \\ 0 & 1 & 0 \\ -\sin(\theta)e^{i\zeta} & 0 & \cos(\theta) \end{pmatrix}$ este unitară şi determinaţi raza ei spectrală.
8. Arătaţi că dacă $P$ este o matrice inversabilă atunci $A \to \|A\| := \|P^{-1}AP\|_\infty$ defineşte o normă matriceală. Este ea subordonată unei norme vectoriale?

# Cursul 3

# Rezolvarea sistemelor de ecuaţii liniare
# Metode Directe (I): Metoda Gauss

Metodele directe de rezolvare a unui sistem liniar constau în aducerea acestuia la o formă echivalentă în care sistemul să fie mai simplu de rezolvat: diagonală, triunghiulară inferior sau triunghiulară superior. Acest lucru se poate face într-un număr finit de paşi, prin intermediul unor operaţii elementare care nu schimbă soluţia sistemului:

- schimbarea ordinii de scriere a două ecuaţii în sistem;
- adunarea unei ecuaţii înmulţită eventual cu un număr nenul la o altă ecuaţie.

## 3.1 Matrici elementare

Operaţii elementare care nu schimbă soluţia unui sistem liniarenumerate mai sus se pot descrie prin intermediul acţiunii unor matrici tip.

- **Matrice de permutare** $P_{ij} \in \mathcal{M}_n(\mathbb{R})$: se obţine din matricea identitate schimbând între

ele liniile $i$ şi $j$:

$$P_{ij} = \begin{pmatrix}
1 & . & . & . & 0 & & & & & & & & \\
. & . & & & . & i & & & & j & & & \\
. & & . & & . & \downarrow & & & & \downarrow & & & \\
. & & & . & . & & & & & & & & \\
0 & . & . & . & 1 & & & & & & & & \\
 & & i & \to & & \mathbf{0} & 0 & . & . & 0 & \mathbf{1} & & \\
 & & & & & 0 & 1 & . & . & 0 & 0 & & \\
 & & & & & . & . & . & & . & . & & \\
 & & & & & . & . & & . & . & . & & \\
 & & & & & . & . & & & . & . & & \\
 & & & & & 0 & 0 & . & . & 1 & 0 & & \\
 & & j & \to & & \mathbf{1} & 0 & . & . & 0 & \mathbf{0} & & \\
 & & & & & & & & & & & 1 & . & . & . & 0 \\
 & & & & & & & & & & & . & . & & & . \\
 & & & & & & & & & & & . & & . & & . \\
 & & & & & & & & & & & . & & & . & . \\
 & & & & & & & & & & & 0 & . & . & . & 1
\end{pmatrix}.$$

i) Dată o matrice $A \in \mathcal{M}_{np}(\mathbb{R})$ avem că $P_{ij}A$ se obţine din matricea $A$ schimbând între ele liniile $i$ şi $j$.

ii) $\det(P_{ij}) = -1$, $P_{ij}^{-1} = P_{ij}$.

- **Matrice** $E_{ij}(a) \in \mathcal{M}_n(\mathbb{R})$: se obţine din matricea identitate adaugând $a$ pe poziţia $(i, j)$.

$$E_{ij}(a) = \begin{pmatrix}
1 & . & . & . & 0 & & & & & & & & \\
. & . & & & . & j & & & & i & & & \\
. & & . & & . & \downarrow & & & & \downarrow & & & \\
. & & & . & . & & & & & & & & \\
0 & . & . & . & 1 & & & & & & & & \\
 & & j & \to & & 1 & . & . & . & 0 & & & \\
 & & & & & . & & & & . & & & \\
 & & & & & . & & & & . & & & \\
 & & & & & . & & & & . & & & \\
 & & i & \to & & \mathbf{a} & . & . & . & 1 & & & \\
 & & & & & & & & & & 1 & . & . & . & 0 \\
 & & & & & & & & & & . & . & & & . \\
 & & & & & & & & & & . & & . & & . \\
 & & & & & & & & & & . & & & . & . \\
 & & & & & & & & & & 0 & . & . & . & 1
\end{pmatrix}.$$

i) Dată o matrice $A \in \mathcal{M}_{np}(\mathbb{R})$ avem că $E_{ij}(a)A$ se obţine din matricea $A$ adunând la linia $i$ linia $j$ înmulţită cu $a$.

ii) $\det(E_{ij}(a)) = 1$, $E_{ij}(a)^{-1} = E_{ij}(-a)$.

## 3.2 Rezolvarea sistemelor triunghiulare

Cel mai uşor de rezolval sistem este unul diagonal:

$$\left\{\begin{array}{rcl} a_{11}x_1 & = & b_1 \\ a_{22}x_2 & = & b_2 \\ a_{33}x_3 & = & b_3 \\ \ddots & & \vdots \\ a_{n-1n-1}x_{n-1} & = & b_{n-1} \\ a_{nn}x_n & = & b_n \end{array}\right. \tag{3.1}$$

a cărui soluţie se obţine imediat:

$$x_i = \frac{b_i}{a_{ii}} \; i = 1, 2, \ldots, n. \tag{3.2}$$

Forma diagonală este prea particulară, de aceea vom analiza alte două forme particulare de sisteme.

### 3.2.1 Sisteme triunghiulare inferior

Un sistem triunghiular inferior este un sistem a cărui matrice este triunghiulară inferior:

$$\left\{\begin{array}{ccccccccccc} a_{11}x_1 & & & & & & & & & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & & & & & & & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & & & & & = & b_3 \\ \vdots & & \vdots & & & \ddots & & & & & \vdots \\ a_{n-11}x_1 & + & a_{n-12}x_2 & + & \ldots & + & a_{n-1n-1}x_{n-1} & & & = & b_{n-1} \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \ldots & + & a_{nn-1}x_{n-1} & + & a_{nn}x_n & = & b_n \end{array}\right. \tag{3.3}$$

Matricea coeficienţilor sistemului (3.3) este

$$A = \begin{pmatrix} a_{11} & 0 & 0 & \ldots & 0 & 0 \\ a_{21} & a_{22} & 0 & \ldots & 0 & 0 \\ a_{31} & a_{32} & a_{33} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-11} & a_{n-12} & a_{n-13} & \ldots & a_{n-1n-1} & 0 \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn-1} & a_{nn} \end{pmatrix}$$

şi este o matrice triunghiulară inferior: $a_{ij} = 0$ dacă $j > i$.

Sistemul (3.3) se poate rezolva imediat, începând cu prima ecuaţie care ne dă prima necunoscută $x_1$, apoi a doua ecuaţie care ne dă a doua necunoscută $x_2$, ş.a.m.d. (rezolvare prin substituţie înainte). Oţinem următoarele formule:

$$\left\{\begin{array}{l} x_1 = \dfrac{b_1}{a_{11}} \\ \\ x_i = \dfrac{1}{a_{ii}} \left( b_i - \displaystyle\sum_{j=1}^{i-1} a_{ij}x_j \right), \quad i = 2, 3, \ldots, n. \end{array}\right. \tag{3.4}$$

Numărul de operaţii necesare găsirii lui $x_i$:

- Adunări: $i-1$
- Înmulţiri: $i-1$
- Împărţiri: 1.

În total avem: $\sum_{i=1}^{n}[1+2(i-1)] = n + 2\frac{n(n-1)}{2} = n^2$ operaţii.

### 3.2.2 Sisteme triunghiulare superior

Un sistem triunghiular superior este un sistem a cărui matrice este triunghiulară superior:

$$(3.5)\qquad \begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1 \\ a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2 \\ a_{33}x_3 + \dots + a_{3n}x_n = b_3 \\ \ddots \quad \vdots \quad \vdots \\ a_{n-1n-1}x_{n-1} + a_{n-1n}x_n = b_{n-1} \\ a_{nn}x_n = b_n \end{cases}$$

Matricea coeficienţilor sistemului (3.5) este

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

şi este o matrice triunghiulară superior: $a_{ij} = 0$ dacă $i > j$.

Sistemul (3.5) se poate rezolva imediat, începând cu ultima ecuaţie care ne dă ultima necunoscută $x_n$, apoi penultima ecuaţie care ne dă penultima necunoscută $x_{n-1}$, ş.a.m.d. (rezolvare prin substituţie înapoi). Oţinem următoarele formule:

$$(3.6)\qquad \begin{cases} x_n = \dfrac{b_n}{a_{nn}} \\ x_i = \dfrac{1}{a_{ii}}\left(b_i - \displaystyle\sum_{j=i+1}^{n} a_{ij}x_j\right), \quad i = n-1, n-2, \dots, 1. \end{cases}$$

Numărul de operaţii necesare găsirii lui $x_i$:

- Adunări: $n-i$
- Înmulţiri: $n-i$

- Împărţiri: 1.

În total avem:

$$\sum_{i=1}^{n}[1+2(n-i)] = n + 2\frac{n(n-1)}{2} =$$

$$= n^2 \text{ operaţii.}$$

## 3.3 Metoda Gauss (Eliminare gaussiană)

Una dintre metodele cele mai vechi de a rezolva un sistem de ecuaţii liniare, folosite şi în prezent, este metoda Gauss, care se mai numeşte şi metoda eliminării gaussiene. Metoda îşi propune să aducă sistemul la forma triunghiulară superior (sau cel puţin la forma scară) într-un număr finit de paşi. Acest lucru se realizează prin eliminarea a câte o necunoscută din ecuaţiile sistemului prin parcurgerea următoarelor etape:

Etapa 1°: Se crează matricea extinsă a sistemului:

$$\widetilde{A}^0 = (A|b).$$

- Se aduce pe poziţia $(1,1)$ un coeficient nenul (eventual prin schimbarea liniei întâi cu o altă linie $i_1$ care are proprietatea că $a^0_{i_1 1} \neq 0$). Acest element se va numi primul pivot. Vom avea eventual o nouă matrice, dar pe care o vom nota la fel pentru simplitate, obţinută prin:

$$\widetilde{A}^0 = P_{i_1 1}\widetilde{A}^0. \tag{3.7}$$

- Apoi se crează zerouri sub pivot (se elimină din toate celelalte ecuaţii necunoscuta $x_1$) prin adunarea liniei întâi înmulţite cu câte un număr la fiecare din celelalte linii ). Operaţiile matriciale care se execută în această etapă sunt:

$$\widetilde{A}^1 = E_{n1}\left(-\frac{a^0_{n1}}{a^0_{11}}\right) E_{n-11}\left(-\frac{a^0_{n-11}}{a^0_{11}}\right)\ldots E_{21}\left(-\frac{a^0_{21}}{a^0_{11}}\right)\widetilde{A}^0 \tag{3.8}$$

La sfârşitul acestei etape se obţine matricea:

$$\widetilde{A}^1 = \begin{pmatrix}
a^1_{11} & a^1_{12} & a^1_{13} & \ldots & a^1_{1k-1} & a^1_{1k} & a^1_{1k+1} & \ldots & a^1_{1n} & a^1_{1n+1} \\
0 & a^1_{22} & a^1_{23} & \ldots & a^1_{2k-1} & a^1_{2k} & a^1_{2k+1} & \ldots & a^1_{2n} & a^1_{2n+1} \\
0 & a^1_{32} & a^1_{33} & \ldots & a^1_{3k-1} & a^1_{3k} & a^1_{3k+1} & \ldots & a^1_{3n} & a^1_{3n+1} \\
\vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\
0 & a^1_{k-12} & a^1_{k-13} & \ldots & a^1_{k-1k-1} & a^1_{k-1k} & a^1_{k-1k+1} & \ldots & a^1_{k-1n} & a^1_{k-1n+1} \\
0 & a^1_{k2} & a^1_{k3} & \ldots & a^1_{kk-1} & a^1_{kk} & a^1_{kk+1} & \ldots & a^1_{kn} & a^1_{kn+1} \\
0 & a^1_{k+12} & a^1_{k+13} & \ldots & a^1_{k+1k-1} & a^1_{k+1k} & a^1_{k+1k+1} & \ldots & a^1_{k+1n} & a^1_{k+11n+1} \\
\vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\
0 & a^1_{n-12} & a^1_{n-13} & \ldots & a^1_{n-1k-1} & a^1_{n-1k} & a^1_{n-1k+1} & \ldots & a^1_{n-1n} & a^1_{n-1n+1} \\
0 & a^1_{n2} & a^1_{n3} & \ldots & a^1_{nk-1} & a^1_{nk} & a^1_{nk+1} & \ldots & a^1_{nn} & a^1_{nn+1}
\end{pmatrix}$$

În continuare vom presupune că am realizat zerouri sub elementele de pe poziţiile $(1,1)$, $(2,2)$,...,$(k-1,k-1)$. Prin urmare la sfârşitul Etapei $(k-1)^\circ$ vom dispune de o matrice de forma

$$\widetilde{A}^{k-1} = \begin{pmatrix} a_{11}^{k-1} & a_{12}^{k-1} & a_{13}^{k-1} & \dots & a_{1k-1}^{k-1} & a_{1k}^{k-1} & a_{1k+1}^{k-1} & \dots & a_{1n}^{k-1} & a_{1n+1}^{k-1} \\ 0 & a_{22}^{k-1} & a_{23}^{k-1} & \dots & a_{2k-1}^{k-1} & a_{2k}^{k-1} & a_{2k+1}^{k-1} & \dots & a_{2n}^{k-1} & a_{2n+1}^{k-1} \\ 0 & 0 & a_{33}^{k-1} & \dots & a_{3k-1}^{k-1} & a_{3k}^{k-1} & a_{3k+1}^{k-1} & \dots & a_{3n}^{k-1} & a_{3n+1}^{k-1} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{k-1k-1}^{k-1} & a_{k-1k}^{k-1} & a_{k-1k+1}^{k-1} & \dots & a_{k-1n}^{k-1} & a_{k-1n+1}^{k-1} \\ 0 & 0 & 0 & \dots & 0 & a_{kk}^{k-1} & a_{kk+1}^{k-1} & \dots & a_{kn}^{k-1} & a_{k}^{k-1,n+1} \\ 0 & 0 & 0 & \dots & 0 & a_{k+1k}^{k-1} & a_{k+1k+1}^{k-1} & \dots & a_{k+1n}^{k-1} & a_{k+1,n+1}^{k-1} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & a_{n-1k}^{k-1} & a_{n-1k+1}^{k-1} & \dots & a_{n-1n}^{k-1} & a_{n-1n+1}^{k-1} \\ 0 & 0 & 0 & \dots & 0 & a_{nk}^{k-1} & a_{nk+1}^{k-1} & \dots & a_{nn}^{k-1} & a_{nn+1}^{k-1} \end{pmatrix}$$

Etapa $k^\circ$: Pornind de la matricea $\widetilde{A}^{k-1}$ von crea zerouri sub poziţia $(k,k)$.

- Se aduce pe poziţia $(k,k)$ un coeficient nenul (eventual prin schimbarea liniei $k$ cu o altă linie $i_k > k$ care are proprietatea că $a_{i_kk}^{k-1} \neq 0$). Acest element se va numi pivotul $k$. Vom avea eventual o nouă matrice, dar pe care o vom nota la fel pentru simplitate, obţinută prin:

$$\widetilde{A}^{k-1} = P_{i_kk}\widetilde{A}^{k-1}. \tag{3.9}$$

- Apoi se crează zerouri sub pivot (se elimină din ecuaţiile $k+1$, ..., $n$ necunoscuta $x_k$) prin adunarea liniei $k$ înmulţite cu câte un număr la fiecare dintre liniile situate sub ea). Operaţiile matriciale care se execută în această etapă sunt:

$$\widetilde{A}^{k} = E_{nk}\left(-\frac{a_{nk}^{k-1}}{a_{kk}^{k-1}}\right) E_{n-1k}\left(-\frac{a_{n-1k}^{k-1}}{a_{kk}^{k-1}}\right) \dots E_{k+1k}\left(-\frac{a_{k+1k}^{k-1}}{a_{kk}^{k-1}}\right) \widetilde{A}^{k-1} \tag{3.10}$$

La sfârşitul Etapei $k^\circ$ vom dispune de o matrice de forma:

$$\widetilde{A}^k = \begin{pmatrix} a_{11}^k & a_{12}^k & a_{13}^k & \dots & a_{1k-1}^k & a_{1k}^k & a_{1k+1}^k & \dots & a_{1n}^k & a_{1n+1}^k \\ 0 & a_{22}^k & a_{23}^k & \dots & a_{2k-1}^k & a_{2k}^k & a_{2k+1}^k & \dots & a_{2n}^k & a_{2n+1}^k \\ 0 & 0 & a_{33}^k & \dots & a_{3k-1}^k & a_{3k}^k & a_{3k+1}^k & \dots & a_{3n}^k & a_{3n+1}^k \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{k-1k-1}^k & a_{k-1k}^k & a_{k-1k+1}^k & \dots & a_{k-1n}^k & a_{k-1n+1}^k \\ 0 & 0 & 0 & \dots & 0 & a_{kk}^k & a_{kk+1}^k & \dots & a_{kn}^k & a_{kn+1}^k \\ 0 & 0 & 0 & \dots & 0 & 0 & a_{k+1k+1}^k & \dots & a_{k+1n}^k & a_{k+1n+1}^k \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & a_{n-1k+1}^k & \dots & a_{n-1n}^k & a_{n-1n+1}^k \\ 0 & 0 & 0 & \dots & 0 & 0 & a_{nk+1}^k & \dots & a_{nn}^k & a_{nn+1}^k \end{pmatrix}$$

Formulele care au permis determinarea matricii $\widetilde{A}^k$ sunt următoarele:

$$a_{ij}^k = a_{ij}^{k-1} - \frac{a_{ik}^{k-1}}{a_{kk}^{k-1}} a_{kj}^{k-1}, \quad k+1 \le i \le n, \quad k \le j \le n+1. \tag{3.11}$$

În continuare vom continua să realizăm zerouri sub elementele de pe poziţiile $(k+1, k+1)$, $\dots$, $(n-2, n-2)$ şi vom ajunge la sfârşitul Etapei $(n-2)^\circ$ să dispune de o matrice de forma:

$$\widetilde{A}^{n-2} = \begin{pmatrix} a_{11}^{n-2} & a_{12}^{n-2} & \dots & a_{1k}^{n-2} & a_{1k+1}^{n-2} & \dots & a_{1n-2}^{n-2} & a_{1n-1}^{n-2} & a_{1n}^{n-2} & a_{1n+1}^{n-2} \\ 0 & a_{22}^{n-2} & \dots & a_{2k}^{n-2} & a_{2k+1}^{n-2} & \dots & a_{2n-2}^{n-2} & a_{2n-1}^{n-2} & a_{2n}^{n-2} & a_{2n+1}^{n-2} \\ 0 & 0 & \dots & a_{3k}^{n-2} & a_{3k+1}^{n-2} & \dots & a_{3n-2}^{n-2} & a_{3n-1}^{n-2} & a_{3n}^{n-2} & a_{3n+1}^{n-2} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & a_{k-1k}^{n-2} & a_{k-1k+1}^{n-2} & \dots & a_{k-1n-2}^{n-2} & a_{k-1n-1}^{n-2} & a_{k-1n}^{n-2} & a_{k-1n+1}^{n-2} \\ 0 & 0 & \dots & a_{kk}^{n-2} & a_{kk+1}^{n-2} & \dots & a_{kn-2}^{n-2} & a_{kn-1}^{n-2} & a_{kn}^{n-2} & a_{kn+1}^{n-2} \\ 0 & 0 & \dots & 0 & a_{k+1k+1}^{n-2} & \dots & a_{k+1n-2}^{n-2} & a_{k+1n-1}^{n-2} & a_{k+1n}^{n-2} & a_{k+1n+1}^{n-2} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & a_{n-2n-2}^{n-2} & a_{n-2n-1}^{n-2} & a_{n-2n}^{n-2} & a_{n-2n+1}^{n-2} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & a_{n-1n-1}^{n-2} & a_{n-1n}^{n-2} & a_{n-1n+1}^{n-2} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & a_{nn-1}^{n-2} & a_{nn}^{n-2} & a_{nn+1}^{n-2} \end{pmatrix}$$

Etapa $(n-1)^\circ$: realizăm zerouri sub elementul de pe poziţiile $(n-1, n-1)$.

- Se aduce pe poziţia $(n-1, n-1)$ un coeficient nenul (eventual prin schimbarea liniei $n-1$ cu linia $i_{n-1} = n$ care are proprietatea că $a^{n-2}_{i_{n-1}n-1} \neq 0$). Acest element se va numi pivotul $n-1$. Vom avea eventual o nouă matrice, dar pe care o vom nota la fel pentru simplitate, obţinută prin:

$$\widetilde{A}^{n-2} = P_{i_{n-1}n-1}\widetilde{A}^{n-2}. \tag{3.12}$$

- Apoi se crează zerouri sub pivot (se elimină din ecuaţia $n$ necunoscuta $x_{n-1}$) prin adunarea liniei $n-1$ înmulţită cu un număr lalinia $n$). Operaţiile matriciale care se execută în această etapă sunt:

$$\widetilde{A}^{n-1} = E_{nn-1}\left(-\frac{a^{n-2}_{nn-1}}{a^{n-2}_{n-1n-1}}\right)\widetilde{A}^{n-2}. \tag{3.13}$$

Vom ajunge la sfârşitul Etapei $(n-1)^\circ$ să dispune de o matrice de forma:

$$\widetilde{A}^{n-1} = \begin{pmatrix} a^{n-1}_{11} & a^{n-1}_{12} & \dots & a^{n-1}_{1k} & a^{n-1}_{1k+1} & \dots & a^{n-1}_{1n-2} & a^{n-1}_{1n-1} & a^{n-1}_{1n} & a^{n-1}_{1} \\ 0 & a^{n-1}_{22} & \dots & a^{n-1}_{2k} & a^{n-1}_{2k+1} & \dots & a^{n-1}_{2n-2} & a^{n-1}_{2n-1} & a^{n-1}_{2n} & a^{n-1}_{2} \\ 0 & 0 & \dots & a^{n-1}_{3k} & a^{n-1}_{3k+1} & \dots & a^{n-1}_{3n-2} & a^{n-1}_{3n-1} & a^{n-1}_{3n} & a^{n-1}_{3} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & a^{n-1}_{k-1k} & a^{n-1}_{k-1k+1} & \dots & a^{n-1}_{k-1n-2} & a^{n-1}_{k-1n-1} & a^{n-1}_{k-1n} & a^{n-1}_{k-1} \\ 0 & 0 & \dots & a^{n-1}_{kk} & a^{n-1}_{kk+1} & \dots & a^{n-1}_{kn-2} & a^{n-1}_{kn-1} & a^{n-1}_{kn} & a^{n-1}_{k} \\ 0 & 0 & \dots & 0 & a^{n-1}_{k+1k+1} & \dots & a^{n-1}_{k+1n-2} & a^{n-1}_{k+1n-1} & a^{n-1}_{k+1n} & a^{n-1}_{k+1} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & a^{n-1}_{n-2n-2} & a^{n-1}_{n-2n} & a^{n-1}_{n-2} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & a^{n-1}_{n-1n-1} & a^{n-1}_{n-1n} & a^{n-1}_{n-1} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & a^{n-1}_{nn} & a^{n-1}_{n} \end{pmatrix}$$

Prin aceste transformări elementare asupra sistemului de ecuaţii liniare se elimină necunoscuta $x_1$ din ecuaţiile $2, 3, \dots, n$, necunoscuta $x_2$ din ecuaţiile $3, 4, \dots, n$ şi aşa mai departe, pentru a obţine un sistem de ecuaţii triunghiular (a cărui matrice asociată va fi o matrice triunghiulară superior).

**Observaţia 3.1.** *Dacă matricea $A$ este inversabilă, deci sistemul iniţial are soluţie unică pentru orice termen liber $b$, atunci la fiecare pas putem găsi pivoţi nenuli:*

- *Pe prima coloană există cel puţin un element nenul, altfel am avea o coloană zero iar matricea nu ar fi inversabilă. Prin urmare găsim un prim pivot $a^0_{i1} \neq 0$.*

- *La etapa $k^\circ$, dacă am avea toate elementele de pe poziţiile $(k,k)$, $(k+1,k)$, $\dots$, $(n,k)$ egale*

*cu zero atunci ar rezulta că*

$$\det \begin{pmatrix} a_{11}^{k-1} & a_{12}^{k-1} & \dots & a_{1k-1}^{k-1} & a_{1k}^{k-1} & a_{1k+1}^{k-1} & \dots & a_{1n}^{k-1} \\ 0 & a_{22}^{k-1} & \dots & a_{2k-1}^{k-1} & a_{2k}^{k-1} & a_{2k+1}^{k-1} & \dots & a_{2n}^{k-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{k-1k-1}^{k-1} & a_{k-1k}^{k-1} & a_{k-1k+1}^{k-1} & \dots & a_{k-1n}^{k-1} \\ 0 & 0 & \dots & 0 & a_{kk}^{k-1} & a_{kk+1}^{k-1} & \dots & a_{kn}^{k-1} \\ 0 & 0 & \dots & 0 & a_{k+1k}^{k-1} & a_{k+1k+1}^{k-1} & \dots & a_{k+1n}^{k-1} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & a_{n-1k}^{k-1} & a_{n-1k+1}^{k-1} & \dots & a_{n-1n}^{k-1} \\ 0 & 0 & \dots & 0 & a_{nk}^{k-1} & a_{nk+1}^{k-1} & \dots & a_{nn}^{k-1} \end{pmatrix} =$$

$$= a_{11}^{k-1} \dots a_{k-1k-1}^{k-1} \begin{pmatrix} a_{kk}^{k-1} & a_{kk+1}^{k-1} & \dots & a_{kn}^{k-1} \\ a_{k+1k}^{k-1} & a_{k+1k+1}^{k-1} & \dots & a_{k+1n}^{k-1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n-1k}^{k-1} & a_{n-1k+1}^{k-1} & \dots & a_{n-1n}^{k-1} \\ a_{nk}^{k-1} & a_{nk+1}^{k-1} & \dots & a_{nn}^{k-1} \end{pmatrix} = 0,$$

*ceea ce ar contrazice inversabilitatea matricii $A$.*

Numărul de operaţii în Etapa $k^\circ$:

- Adunări: $(n-k)(n-k+1)$
- Înmulţiri: $(n-k)(n-k+1)$
- Împărţiri: $n-k$.

În total avem:

$$\sum_{k=1}^{n-1} [(n-k) + 2(n-k)(n-k+1)] = 3\sum_{k=1}^{n-1}(n-k) + 2\sum_{k=1}^{n-1}(n-k)^2 =$$

$$= 3\frac{n(n-1)}{2} + 2\frac{(n-1)n(2n-1)}{6} =$$

$$= \mathcal{O}\left(\frac{2}{3}n^3\right) \text{ operaţii.}$$

## 3.4 Forma scară a unei matrici

Metoda lui Gauss se poate aplica oricărui tip de sistem de ecuaţii liniare dar nu vom ajunge la forma triunghiulară, ci la forma scară.

**Definiţia 3.1.** *Se spune că o matrice este în formă de scară dacă:*

*(i) Dacă linia $k$ nu constă în întregime din zerouri, numărul de elemente egale cu zero de la începutul liniei $k+1$ este mai mare decât numărul de elemente egale cu zero de la începutul liniei $k$.*

*(ii) Dacă există linii ale căror elemente sunt toate zero, ele sunt sub liniile care au elemente nenule.*

Un exemplu schematic de matrice care este în formă de scară este următorul (elementele marcate cu $*$ sunt nenule):

$$\begin{pmatrix} 0 & * & * & 0 & * & * & 0 \\ 0 & 0 & 0 & * & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Pornind de la orice matrice putem ajunge la forma scară aplicând metoda eliminării gaussiene. Atunci când pe o coloană nu găsim un pivot nenul, trecem la coloana următoare. Numărul de pivoţi nenuli (sau, echivalent, de linii nenule în forma scară) este egal cu rangul matricii iniţiale.

# Cursul 4

# Metode Directe (II): Factorizări

## 4.1 Factorizarea LU

Fie $A \in \mathcal{M}_n(\mathbb{R})$ o matrice pătratică. Vom presupune că $A$ este inversabilă şi că procesul de eliminare gaussiană se poate face fără a fi nevoie de inversare de linii. Reluând procedeul de eliminare gaussiană avem că:

$$\underbrace{E_{nn-1}\left(-\frac{a^{n-2}_{nn-1}}{a^{n-2}_{n-1n-1}}\right)}_{\text{Etapa } n-1} \ \dots \ \underbrace{E_{n1}\left(-\frac{a^0_{n1}}{a^0_{11}}\right) E_{n-11}\left(-\frac{a^0_{n-11}}{a^0_{11}}\right) \dots E_{21}\left(-\frac{a^0_{21}}{a^0_{11}}\right)}_{\text{Etapa } 1} A = U,$$

unde $U$ este o matrice triunghiulară superior. Vom obţine că:

$$A = E_{21}\left(\frac{a^0_{21}}{a^0_{11}}\right) \dots E_{n-11}\left(\frac{a^0_{n-11}}{a^0_{11}}\right) E_{n1}\left(\frac{a^0_{n1}}{a^0_{11}}\right) \ \dots \ E_{nn-1}\left(\frac{a^{n-2}_{nn-1}}{a^{n-2}_{n-1n-1}}\right) U.$$

Ţinând cont de regula de în mulţire la stânga cu matrici de forma $E_{ij}(a)$, vom obţine că

$$E_{21}\left(\frac{a^0_{21}}{a^0_{11}}\right) \dots E_{n-11}\left(\frac{a^0_{n-11}}{a^0_{11}}\right) E_{n1}\left(\frac{a^0_{n1}}{a^0_{11}}\right) \ \dots \ E_{nn-1}\left(\frac{a^{n-2}_{nn-1}}{a^{n-2}_{n-1n-1}}\right) =$$

$$= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{a^0_{21}}{a^0_{11}} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{a^0_{n-11}}{a^0_{11}} & \frac{a^1_{n-12}}{a^1_{22}} & \dots & 1 & 0 \\ \frac{a^0_{n1}}{a^0_{11}} & \frac{a^1_{n2}}{a^1_{22}} & \dots & \frac{a^{n-2}_{nn-1}}{a^{n-2}_{n-1n-1}} & 1 \end{pmatrix} = L,$$

unde $L$ este o matrice triunghiulară inferior.

Astfel, obţinem că matricea $A$ se poate scrie sub forma unui produs dintre o matrice triunghiulară superior şi o matrice triunghiulară inferior: $A = LU$. Să observăm că în acest produs $U$ este matricea obţinută în urma procesului de eliminare gaussiană iar $L$ se scrie imediat pe baza indicaţiilor oferite de matricile $E_{ij}(a)$ folosite în timpul procesului de eliminare gaussiană.

Dacă $A = (a_{ij})_{1\le i,j\le n} \in \mathcal{M}_n(\mathbb{R})$, vom introduce următoarele notaţii:

$$\Delta_1 = (a_{11}), \quad \Delta_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \dots \quad \Delta_k = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \dots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix}, \quad \dots$$

$$\Delta_n = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = A.$$

**Teorema 4.1.** *Fie $A \in \mathcal{M}_n(\mathbb{R})$ o matrice cu proprietatea că*

$$\det(\Delta_k) \neq 0 \quad (1 \le k \le n). \tag{4.1}$$

*Există o unică matrice triunghiulară superior $U$ şi o unică matrice triunghiulară inferior cu 1 pe diagonala principală $L$ cu proprietatea că*

$$A = LU \quad (\textit{Factorizare } LU). \tag{4.2}$$

*Proof.* Ţinând cont de ceea ce am obţinut mai ramâne de arătat:

1. *Posibilitatea eliminării gaussiene fără permutări de linii.* Acest lucru are loc dacă, la fiecare etapă a eliminării gaussiene, pe poziţia pivotului se găseşte un element nenul. Arătăm că (4.1) asigură acest lucru.

   - La Etapa 1 pe poziţia $(1,1)$ se găseşte elementul $a_{11}$, care este nenul deoarece $\det(\Delta_1) \neq 0$. Astfel, la prima etapă nu este nevoie de inversare de linii.
   - În urma eliminării gaussiene, după Etapa $k-1$, se obţine că $\det(\Delta_k) = a_{11}^{k-1} \dots a_{kk}^{k-1}$ care, fiind nenul, implică faptul că $a_{kk}^{k-1} \neq 0$. Prin urmare, la Etapa $k$ pe poziţia $(k,k)$ se găseşte elementul nenul $a_{kk}^{k-1}$ şi prin urmare nu este nevoie de inversare de linii.

2. *Unicitarea factorizării.* Presupunem că $U_1$, $U_2$ sunt două matrici triunghiulare superior şi $L_1$, $L_2$ sunt două matrici triunghiulare inferior cu 1 pe diagonala principală astfel ca

   $$A = L_1U_1 = L_2U_2.$$

   Din relaţia de mai sus deducem că

   $$L_2^{-1}L_1 = U_2U_1^{-1}.$$

   Ţinând cont că $L_2^{-1}L_1$ este o matrice triunghiulară inferior cu 1 pe diagonala principală iar $U_2U_1^{-1}$ este o matrice triunghiulară superior, rezulă că unica posibilitate este:

   $$L_2^{-1}L_1 = U_2U_1^{-1} = I,$$

   ceea ce probează unicitatea factorizării.

□

**Observaţia 4.1.** *Factorizarea $LU$ poate fi folosită la rezolvarea sistemelor de ecuaţii liniare în următorul mod:*

$$Ax = b \Leftrightarrow (LU)x = b \Leftrightarrow \begin{cases} Ly = b \\ Ux = y. \end{cases}$$

*Conform acestui algoritm, dacă dispunem de factorizarea $LU$ a matricii sistemului, rezolvarea acestuia se reduce la rezolvarea a două sisteme triunghiulare, unul inferior si altul superior.*

*Să mai menţionăm că această abordare este foarte eficientă dacă avem de rezolvat mai multe sisteme care au aceeaşi matrice dar termeni liberi diferiţi. În acest caz factorizarea se face o singură dată şi este folosită de câte ori este nevoie.*

## 4.2 Factorizare Cholesky

În anumite cazuri factorizarea se poate face mai simplu (involucrând doar o matrice) şi cu un număr mai mic de operaţii decât cel necesar prin eliminarea gaussiană. Vom presupune în cele ce urmează că $A \in \mathcal{M}_n(\mathbb{R})$ este o matrice definită pozitiv. Reamintim că o matrice pătratică $A \in \mathcal{M}_n(\mathbb{C})$ este definită pozitiv dacă este hermitică şi $(Av, v) > 0$ pentru orice $v \in \mathbb{C}^n \setminus \{0\}$. Pentru început să notăm următoarea proprietate.

**Propoziţia 4.1.** *Fie $A \in \mathcal{M}_n(\mathbb{C})$ o matrice definită pozitiv. Atunci matricile $\Delta_k$ sunt definite pozitiv pentru fiecare $k \in \{1, \ldots, n\}$.*

*Proof.* Evident, fiecare matrice $\Delta_k$ este hermitică. Pentru orice $v \in \mathbb{C}^k \setminus \{0\}$ avem că

$$v^* \Delta_k v = \widetilde{v}^* A \widetilde{v} > 0,$$

unde $\widetilde{v} \in \mathbb{C}^n$ se obţine adăugând $n - k$ componente egale cu zero la $v$. □

**Observaţia 4.2.** *Matricile $\Delta_k$ fiind definite pozitiv au determinantul nenul.*

**Teorema 4.2.** *Fie $A \in \mathcal{M}_n(\mathbb{R})$ o matrice definită pozitiv. Există o unică matrice triunghiulară inferior cu elemente pozitive pe diagonala principală $B$ cu proprietatea că*

$$A = BB^t \quad (\textit{Factorizare Cholesky}). \tag{4.3}$$

*Proof.*

1. *Existenţa factorizării.* Conform Teoremei 4.1 există o matrice triunghiulară superior $U$ şi o unică matrice triunghiulară inferior cu 1 pe diagonala principală $L$ cu proprietatea că $A = LU$. Deoarece matricile $\Delta_k$ sunt definite pozitiv, rezultă că elementele de pe diagonala principală a lui $U$ sunt pozitive. Vom scrie $U = D^2\widetilde{U}$, unde $\widetilde{U}$ este o matrice triunghiulară superior cu 1 pe diagonala principală iar $D$ este o matrice diagonală cu elemente pozitive pe diagonala principală.

   Cum matricea $A$ este simetrică şi $A = A^t = \widetilde{U}^t D^2 L^t$, din unicitatea factorizării $LU$ rezultă că $L = \widetilde{U}^t$. Obţinem că

   $$A = LD^2\widetilde{U} = (LD)(\widetilde{U}^t D)^t = (LD)(LD)^t,$$

   adică exact (4.3) cu $B = LD$.

2. *Unicitarea factorizării.* Presupunem că $B_1$, $B_2$ sunt două matrici triunghiulare inferior cu elemente pozitive pe diagonala principală astfel ca

$$A = B_1 B_1^t = B_2 B_2^t.$$

Din relaţia de mai sus deducem că

$$B_2^{-1} B_1 = B_2^t (B_1^t)^{-1}.$$

Ţinând cont că $B_2^{-1} B_1$ este o matrice triunghiulară inferior iar $B_2^t (B_1^t)^{-1}$ este o matrice triunghiulară superior, rezulă că unica posibilitate este:

$$B_2^{-1} B_1 = B_2^t (B_1^t)^{-1} = D,$$

unde $D$ este o matrice diagonală. Cum avem că

$$I = B_2^{-1} B_1 B_1^t (B_2^t)^{-1} = B_2^{-1} B_1 (B_2^{-1} B_1)^t = D^2,$$

iar $D$ are elemente pozitive pe diagonala principală, rezultă că $D = I$ şi se obţine unicitatea factorizării.

□

Deducerea matricii $B$: Din relaţia $A = BB^t$, deducem că

$$a_{ij} = \sum_{k=1}^{\min\{i,j\}} b_{ik} b_{jk}. \tag{4.4}$$

- Pentru $i = 1$ obţinem că

$$b_{11} = \sqrt{a_{11}}, \qquad b_{j1} = \frac{a_{1j}}{b_{11}} \quad (2 \le j \le n), \tag{4.5}$$

ceea ce reprezintă prima coloană a lui $B$.

- Presupunem calculate elementele de pe coloanele $1, \ldots, i-1$ ale lui $B$ şi le determinăm pe cele de pe coloana $i$.

$$b_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} b_{ik}^2}, \qquad b_{ji} = \frac{1}{b_{ii}} \left( a_{ij} - \sum_{k=1}^{i-1} b_{ik} b_{jk} \right) \quad (i+1 \le j \le n). \tag{4.6}$$

Numărul de operaţii pentru coloana $i$:

- Adunări: $(i-1)(n-i+1)$
- Înmulţiri: $(i-1)(n-i+1)$
- Împărţiri: $n-i$
- Radicali: 1.

În total avem:

$$\sum_{i=1}^{n}[1+(n-i)+2(i-1)(n-i+1)] = n+(2n+1)\sum_{i=1}^{n}(i-1)-2\sum_{i=1}^{n}(i-1)^2 =$$

$$= n+\frac{(2n+1)n(n-1)}{2}-2\frac{(n-1)n(2n-1)}{6} =$$

$$=\mathcal{O}\left(\frac{1}{3}n^3\right) \text{ operaţii.}$$

## 4.3 Exerciţii

1. Fie
$$A=\begin{pmatrix} 4 & 1 & 2 & 0\\ 0 & 3 & -1 & 1\\ 4 & 1 & -7 & 1\\ 8 & 5 & -6 & 20 \end{pmatrix} \quad \text{şi} \quad b=\begin{pmatrix} 7\\ 3\\ -1\\ 27 \end{pmatrix}.$$
Rezolvaţi sistemul $Ax=b$ folosind factorizarea LU a matricii $A$.

2. Fie
$$A=\begin{pmatrix} 1 & 3 & 1 & 0\\ 3 & 10 & 2 & -1\\ 1 & 2 & 3 & 3\\ 0 & -1 & 3 & 6 \end{pmatrix} \quad \text{şi} \quad b=\begin{pmatrix} 5\\ 14\\ 9\\ 8 \end{pmatrix}.$$
Rezolvaţi sistemul $Ax=b$ folosind factorizarea Cholesky a matricii $A$.

3. Fie $t\in\mathbb{R}\setminus\{0\}$ şi
$$A=\begin{pmatrix} 1 & 1 & 1\\ t & 2t & 2\\ t+1 & 0 & 2t \end{pmatrix} \quad \text{şi} \quad b=\begin{pmatrix} 1\\ 0\\ 1 \end{pmatrix}.$$
Arătaţi că există o factorizare $LU$ a matricii $A$ şi rezolvaţi sistemul $Ax=b$ folosind această factorizare.

4. Fie $A\in\mathcal{M}_n(\mathbb{R})$ o matrice care verifică ipotezele din Teorema factorizării LU. Dacă $A$ este simetrică arătaţi că există o unică matrice triunghiulară inferior cu 1 pe diagonala principală $L$ şi o unică matrice diagonală $D$ astfel încât $A=LDL^t$.

5. Considerăm aplicaţia $S:\mathbb{R}^2\to\mathbb{R}^2$ definită prin
$$S\begin{pmatrix} v_1\\ v_2 \end{pmatrix}=\begin{pmatrix} v_1\\ -v_2 \end{pmatrix}.$$
   (a) Determinaţi matricea $A\in\mathcal{M}_2(\mathbb{R})$ astfel ca $Sv=Av$, pentru orice $v\in\mathbb{R}^2$.
   (b) Arătaţi că $S$ este o aplicaţie liniară. Ce reprezintă geometric transformarea $S$?
   (c) Determinaţi valorile şi vectorii proprii ai matricii $A$. Care este semnificaţia lor?

6. Fie $A\in\mathcal{M}_n(\mathbb{C})$ o matrice inversabilă şi $E=I-\frac{1}{\mathrm{Tr}(AA^*)}AA^*$. Demonstraţi că $\rho(E)<1$.
   ($\mathrm{Tr}(M)$ este urma matricii $M$, adică suma elementelor de pe diagonala principală a lui $M$.)

# Cursul 5

# Metode iterative pentru rezolvarea sistemelor de ecuaţiilor liniare

Considerăm sistemul de ecuaţii liniare:

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & ... & + & a_{1n}x_n & = b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & ... & + & a_{2n}x_n & = b_2 \\ ..... & ... & ..... & ... & ..... & ... & ..... & ..... \\ a_{n1}x_1 & + & a_{n2}x_2 & + & ... & + & a_{nn}x_n & = b_n \end{array}\right. \tag{5.1}$$

pe care îl scriem în formă matriceală astfel:

$$Ax = b, \tag{5.2}$$

unde $A \in \mathcal{M}_n(\mathbb{R})$ este matricea sistemului, $b \in \mathbb{R}^n$ este termenul liber iar $x \in \mathbb{R}^n$ este vectorul necunoscută. În acest capitol vom presupune că sistemul (5.2) are o unică soluţie $x^* \in \mathbb{R}^n$.

Ideea grupului de metode următor este aceea de a construi un şir de vectori din $\mathbb{R}^n$ (sau din $\mathbb{C}^n$): $x^0$, $x^1$, $x^2$,...care să convergă la $x^*$. Astfel vom avea aproximări din ce în ce mai bune pentru $x^*$.

Spre deosebire de metodele directe, care îşi propuneau să găsească soluţia exactă a sistemului într-un număr finit de paşi, metodele iterative următoare generează un şir infinit de de aproximări. Dacă avem asigurată convergenţa şirului, ne putem opri la un anumit pas, să zicem $k$, şi să luăm $x^k$ ca aproximare a lui $x^*$. În multe situaţii practice, acest lucru este la fel de satisfăcător precum găsirea soluţiei exacte. Cu atât mai mult cu cât, de cele mai multe ori, cu metodele directe nu obţinem chiar soluţia exactă datorită erorilor inerente de rotunjire a numerelor.

## 5.1 Metode iterative

Vom scrie sistemul (5.2) în forma echivalentă:

$$x = Bx + c. \tag{5.3}$$

Subliniem faptul că sistemul (5.3) este echivalent cu (5.2), adică ambele au aceeaşi soluţie unică $x^*$. Prin urmare, avem că

$$x^* = Bx^* + c. \tag{5.4}$$

**Observaţia 5.1.** *Prin scrierea sistemului* (5.2) *în forma* (5.3) *am transformat problema iniţială într-una de găsire a punctului fix al aplicaţiei*

$$x \in \mathbb{C}^n \to Bx + c.$$

*Pentru aceasta vom aplica metoda aproximaţiilor succesive, învăţată în anul 2 la ecuaţii neliniare.*

Dat $x^0 \in \mathbb{C}^n$, construim şirul de vectori $(x^k)_{k\geq 0}$ definit prin recurenţă astfel:

$$x^k = Bx^{k-1} + c \quad (k \geq 1). \tag{5.5}$$

Scrierea (5.3), împreună cu modalitatea de construcţie (5.5) a şirului $(x^k)_{k\geq 0}$, formează o *metodă iterativă*. Ceea ce ne interesează este dacă şirul $(x^k)_{k\geq 0}$ ne dă aproximări din ce în ce mai bune pentru $x^*$. Vom începe cu următoarea definiţie:

**Definiţia 5.1.** *Metoda iterativă* (5.5) *se numeşte convergentă dacă, pentru orice* $x^0 \in \mathbb{C}^n$*, şirul* $(x^k)_{k\geq 0}$ *dat de* (5.5) *este convergent la* $x^*$*.*

Reamintim că un şir $(x^k)_{k\geq 0} \subset \mathbb{C}^n$ este convergent la $x^* \in \mathbb{C}^n$ dacă, într-o normă vectorială $\| \ \|$ din $\mathbb{C}^n$, are loc

$$\lim_{k\to\infty} \|x^* - x^k\| = 0.$$

Următoarea teoremă ne dă o condiţie necesară şi suficientă pentru convergenţa unei metode iterative.

**Teorema 5.1.** *Metoda iterativă* (5.5) *este convergentă dacă şi numai dacă raza spectrală a matricii* $B$ *este subunitară:*

$$\rho(B) < 1. \tag{5.6}$$

*Proof.* Fie $x^0 \in \mathbb{C}^n$. Din (5.4) şi (5.5) deducem că

$$x^* - x^k = B(x^* - x^{k-1}).$$

Repetând argumentul pentru $x^* - x^{k-1}$, $x^* - x^{k-2}$,...,$x^* - x^1$, obţinem că

$$x^* - x^k = B^k(x^* - x^0). \tag{5.7}$$

Vom demonstra acum cele două implicaţii.
"$\Rightarrow$" Presupunem, prin reducere la absurd, că $\rho(B) \geq 1$. Reamintim că

$$\rho(B) = \max\{|\lambda| \ : \ \lambda \text{ valoare proprie a lui } B\}.$$

Prin urmare există $\lambda \in \mathbb{C}$, valoare proprie a lui $B$ cu $|\lambda| \geq 1$. Fie $v \in \mathbb{C}^n \setminus \{0\}$ un vector propriu corespunzător lui $\lambda$. Luând $x^0 = x^* - v$ şi folosind (5.7) obţinem că

$$x^* - x^k = B^k(x^* - x^* + v) = B^k v = \lambda^k v,$$

şi prin urmare

$$\|x^* - x^k\| = |\lambda|^k \|v\|.$$

Metoda iterativă fiind convergentă, membrul din stânga tinde la zero. În schimb, cum $|\lambda| \geq 1$, membrul din dreapta tinde sau la $\|v\| \neq 0$ (dacă $|\lambda| = 1$) sau la infinit (dacă $|\lambda| > 1$). Am obţinut o contradicţie şi deci $\rho(B) < 1$.

"$\Leftarrow$" Fie $\varepsilon > 0$ astfel ca $\rho(B) + \varepsilon < 1$. Am văzut în cursul introductiv că există o normă matriceală subordonată unei norme vectoriale (le vom nota pe ambele cu $\| \ \|_\varepsilon$) astfel ca

$$\|B\|_\varepsilon < \rho(B) + \varepsilon < 1.$$

Folosind (5.7) obţinem că

$$\|x^* - x^k\|_\varepsilon = \|B^k(x^* - x^0)\|_\varepsilon \leq \|B^k\|_\varepsilon \|v\|_\varepsilon \leq \|B\|_\varepsilon^k \|v\|_\varepsilon.$$

Ţinând cont de faptul că $\|B\|_\varepsilon < 1$ şi trecând la limită în inegalitatea de mai sus, rezultă că

$$\lim_{k\to\infty} \|x^* - x^k\|_\varepsilon = 0.$$

Cum acest lucru are loc indiferent de alegerea lui $x^0$, rezultă că metoda iterativă este convergentă şi demonstraţia se încheie. □

Teorema de mai sus ne dă o condiţie necesară şi suficientă pentru convergenţa unei metode iterative dar este greu de verificat în practică. Determinarea razei spectrale presupune cunoaşterea spectrului matricii iar acest lucru este, în general, dificil. Următoarea teoremă ne oferă o condiţie doar suficientă pentru convergenţa unei metode iterative, dar care este mai uşor de verificat.

**Teorema 5.2.** *Metoda iterativă* (5.5) *este convergentă dacă există o normă matricială subordonată unei norme vectoriale, notate ambele cu* $\| \ \|$*, astfel încât norma matricii* $B$ *este subunitară:*

$$\|B\| < 1. \tag{5.8}$$

*În acest caz, au loc estimările:*

$$\|x^* - x^k\| \leq \frac{\|B\|}{1 - \|B\|} \|x^k - x^{k-1}\| \quad (k \geq 1), \tag{5.9}$$

$$\|x^* - x^k\| \leq \frac{\|B\|^k}{1 - \|B\|} \|x^1 - x^0\| \quad (k \geq 1). \tag{5.10}$$

*Proof.* Cum ştim dintr-o teoremă din cursul introductiv,

$$\rho(B) \leq \|B\|,$$

şi prin urmare convergenţa metodei iterative (5.5) rezultă din Teorema 5.1.

Pentru a demonstra relaţiile (5.9)-(5.10) observăm mai întâi că

$$\|x^{k+1} - x^k\| \leq \begin{cases} \|B\| \, \|x^k - x^{k-1}\| \\ \\ \|B\|^k \, \|x^1 - x^0\| \end{cases} \quad (k \geq 1). \tag{5.11}$$

Cum avem că

$$\|x^{k+p+1}-x^k\| \le \|x^{k+p+1}-x^{k+p}\| + \|x^{k+p}-x^{k+p-1}\| + ... + \|x^{k+1}-x^k\|,$$

aplicând relaţiile (5.11), avem că

$$\|x^{k+p+1}-x^k\| \le \begin{cases} \left(\|B\|^{p+1} + \|B\|^p + ... + \|B\|\right) \|x^k - x^{k-1}\|, \\ \left(\|B\|^{k+p} + \|B\|^{k+p-1} + ... + \|B\|^k\right) \|x^1 - x^0\|. \end{cases} \tag{5.12}$$

Facând $p$ să tindă la infinit în relaţiile (5.12) şi ştiind că

$$\lim_{p\to\infty} \left(\|B\|^p + \|B\|^{p-1} + ... + \|B\| + 1\right) = \frac{1}{1-\|B\|},$$

obţinem cele două inegalităţi (5.9)-(5.10). □

**Observaţia 5.2.** *Teorema anterioră este similară celei din anul 2 pentru analiza metodei aproximaţiilor succesive la ecuaţii neliniare. După cum am mai spus, de fapt avem aceeaşi metodă, aplicată de data aceasta funcţiei vectoriale*

$$x \in \mathbb{C}^n \to Bx + c,$$

*al cărei punct fix este $x^*$. Relaţiile* (5.9)-(5.10) *sunt similare celor învăţate în anul 2, unde în loc de $\|B\|$ apărea constanta de contracţie $q \in (0,1)$.*

*Există însă şi o diferenţă importantă: atunci când avem convergenţă, o avem pentru orice aproximaţie iniţială în $\mathbb{C}^n$, nu doar local în jurul lui $x^*$.*

**Observaţia 5.3.** *Relaţia* (5.9) *ne oferă un criteriu de oprire a iteraţiilor într-un program: dat $\varepsilon > 0$, vom calcula termenii şirului $(x^k)_{k\ge 0}$ până când*

$$\|x^k - x^{k-1}\| \le \varepsilon.$$

*Atunci, conform lui* (5.9)*, $x^k$ va fi "aproape" de $x^*$. Mai precis,*

$$\|x^* - x^k\| \le \frac{\|B\|}{1-\|B\|}\varepsilon.$$

**Observaţia 5.4.** *Relaţia* (5.10) *ne oferă şi posibilitatea de a estima numărul de iteraţii necesar atingerii unei anumite precizii: dat $\varepsilon > 0$, vom că*

$$\|x^* - x^k\| \le \varepsilon,$$

*atunci când*

$$\frac{\|B\|^k}{1-\|B\|}\|x^1 - x^0\| \le \varepsilon,$$

*sau, echivalent, atunci când*

$$k \ge \log_{\|B\|}\left(\frac{1-\|B\|}{\|x^1-x^0\|}\varepsilon\right).$$

# Cursul 6

# Metodele iterative Jacobi şi Gauss-Seidel

Am văzut în cursul precedent că o metodă iterativă pentru aproximarea soluţiei unice $x^* \in \mathbb{R}^n$ a unui sistem de ecuaţii liniare în forma matriceală

$$Ax = b, \tag{6.1}$$

constă în scrierea sistemului (6.1) într-o forma echivalentă de tipul

$$x = Bx + c, \tag{6.2}$$

şi construirea unui şir de vectori, prin recurenţă, astfel

$$\begin{cases} x^0 \text{ dat}, \\ x^k = Bx^{k-1} + c \quad (k \geq 1). \end{cases} \tag{6.3}$$

Am mai văzut că metoda iterativă (6.3) este convergentă (adică şirul $(x^k)_{k\geq 0}$ dat de (6.3) converge la $x^*$, unica soluţie a lui (6.1), pentru orice $x^0 \in \mathbb{C}^n$) dacă cel puţin una dintre condiţiile următoare este verificată:

1. $\rho(B) < 1$ (condiţie necesară şi suficientă);
2. Există o normă matriceală subordonată unei norme verctoriale (ambele notate la fel, $\|\cdot\|$) astfel încât $\|B\| < 1$ (condiţie suficientă dar nu necesară);

În acest curs vom vedea două modalităţi de scriere a sistemului (6.1) într-o formă (6.2) şi vom da criterii uşor de verificat pentru convergenţa metodelor iterative rezultate. Începem prin a introduce două categorii importante de matrici.

**Definiţia 6.1.** *O matrice $A = (a_{ij})_{1\leq i,j\leq n} \in \mathcal{M}_n(\mathbb{C})$ este strict diagonal dominantă dacă*

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \qquad (1 \leq i \leq n). \tag{6.4}$$

*Dacă în (6.4) în loc de $>$ avem $\geq$, atunci matricea $A$ se numeşte diagonal dominantă.*

Să observăm că o matrice este strict diagonal dominantă (vom folosi abrevierea s.d.d.), dacă modulul fiecărui element de pe diagonala principală este strict mai mare decât suma modulelor celorlaltor elemente de pe aceeaşi linie cu el. De exemplu, dacă

$$A = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 3 & -1 \\ -2 & 4 & 7 \end{pmatrix},$$

atunci avem că

$$\begin{aligned} |a_{11}| = 2 > |a_{12}| + |a_{13}| = 1 + 0 = 1 \\ |a_{22}| = 3 > |a_{21}| + |a_{23}| = 1 + 1 = 2 \\ |a_{33}| = 7 > |a_{31}| + |a_{32}| = 2 + 4 = 6, \end{aligned}$$

şi prin urmare $A$ este s.d.d.

Pe de altă parte, fiind dată o matrice $A = (a_{ij})_{1 \le i,j \le n} \in \mathcal{M}_n(\mathbb{C})$, vom folosi următoarea descompunere a ei:

$$A = -E + D - F,$$

unde $D$ este partea diagonală a lui $A$, $-E$ este partea triunghiulară inferior a lui $A$ şi $-F$ este partea triunghiulară superior a lui $A$:

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n-1} & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n-1} & a_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n-11} & a_{n-12} & a_{n-13} & \dots & a_{n-1n-1} & a_{n-1n} \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn-1} & a_{nn} \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & 0 & \dots & 0 & 0 \\ a_{31} & a_{32} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n-11} & a_{n-12} & a_{n-13} & \dots & 0 & 0 \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn-1} & 0 \end{pmatrix}}_{-E} +$$

$$+ \underbrace{\begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 & 0 \\ 0 & a_{22} & 0 & \dots & 0 & 0 \\ 0 & 0 & a_{33} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{n-1n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & a_{nn} \end{pmatrix}}_{D} + \underbrace{\begin{pmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n-1} & a_{1n} \\ 0 & 0 & a_{23} & \dots & a_{2n-1} & a_{2n} \\ 0 & 0 & 0 & \dots & a_{3n-1} & a_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & a_{n-1n} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}}_{-F}$$

Trecem acum la prezentarea celor două metode iterative care fac subiectul cursului de azi.

## 6.1 Metoda Jacobi

Această metodă tranformă sistemul $Ax = b$ într-o formă echivalentă, folosind descompunerea $A = -E + D - F$ de mai sus, astfel:

$$Ax = b \Leftrightarrow (-E + D - F)x = b \Leftrightarrow Dx = (E + F)x + b \Leftrightarrow x = D^{-1}(E + F)x + D^{-1}b.$$

Prin urmare, notând $B_J = D^{-1}(E + F)$ şi $c_J = D^{-1}b$, avem că (6.1) se scrie în forma echivalentă

$$x = B_J x + c_J. \tag{6.5}$$

Această formă ne permite să definim un şir recurent

$$\begin{cases} x^0 \text{ dat}, \\ x^k = B_J x^{k-1} + c_J \quad (k \geq 1). \end{cases} \tag{6.6}$$

Avem prin urmare o metodă iterativă. Ea poartă numele de *metoda iterativă Jacobi*. Fiind dat $x^0 \in \mathbb{R}^n$, calculăm la fiecare iteraţie $k \geq 1$ un vector $x^k$ ale cărui componente sunt date de formula

$$x_i^k = \frac{1}{a_{ii}} \left( b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j^{k-1} \right) \qquad (1 \leq i \leq n), \tag{6.7}$$

sau, mai desfăşurat,

$$\begin{cases} x_1^k = \dfrac{1}{a_{11}} \left( b_1 - \displaystyle\sum_{j=2}^{n} a_{1j} x_j^{k-1} \right) \\ x_i^k = \dfrac{1}{a_{ii}} \left( b_i - \displaystyle\sum_{j=1}^{i-1} a_{ij} x_j^{k-1} - \sum_{j=i+1}^{n} a_{ij} x_j^{k-1} \right) \qquad (2 \leq i \leq n-1) \\ x_n^k = \dfrac{1}{a_{nn}} \left( b_n - \displaystyle\sum_{j=1}^{n-1} a_{nj} x_j^{k-1} \right). \end{cases} \tag{6.8}$$

**Observaţia 6.1.** *Pentru a putea defini şirul $(x^k)_{k \geq 0}$ prin (6.6) trebuie ca $a_{ii} \neq 0$, pentru $i \in \{1, 2, .., n\}$.*

Avem următorul rezultat privind convergenţa metodei Jacobi.

**Teorema 6.1.** *Dacă $A$ este o matrice strict diagonal dominantă, atunci metoda iterativă Jacobi (6.6) este convergentă.*

*Proof.* Vom arăta că $\|B_J\|_\infty < 1$, ceea ce va implica convergenţa metodei Jacobi conform criteriului suficient 2. Avem că

$$\|B_J\|_\infty = \max_{1 \leq i \leq n} \left\{ \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|}{|a_{ii}|} \right\} = \max_{1 \leq i \leq n} \left\{ \frac{1}{|a_{ii}|} \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right\}.$$

Cum matricea $A$ este s.d.d., rezultă că

$$1 > \frac{1}{|a_{ii}|} \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \qquad (1 \leq i \leq n),$$

şi prin urmare $\|B_J\|_\infty < 1$. □

**Observaţia 6.2.** *Dacă matricea $A$ este s.d.d. atunci*

$$|a_{ii}| > 0 \qquad (1 \leq i \leq n),$$

*ceea ce arată că metoda Jacobi este bine definită.*

## 6.2 Metoda Gauss-Seidel

Această metodă tranformă sistemul $Ax = b$ într-o formă echivalentă, folosind descompunerea $A = -E + D - F$ de mai sus, astfel:

$$Ax = b \Leftrightarrow (-E + D - F)x = b \Leftrightarrow (-E + D)x = Fx + b \Leftrightarrow x = (-E + D)^{-1}Fx + (-E + D)^{-1}b.$$

Prin urmare, notând $B_G = (-E + D)^{-1}F$ şi $c_G = (-E + D)^{-1}b$, avem că (6.1) se scrie poate în forma echivalentă

$$x = B_G x + c_G. \tag{6.9}$$

Acum, definim şirul recurent

$$\begin{cases} x^0 \text{ dat}, \\ x^k = B_G x^{k-1} + c_G \quad (k \geq 1). \end{cases} \tag{6.10}$$

Avem prin urmare o metodă iterativă care poartă numele de *metoda iterativă Gauss-Seidel*. Fiind dat $x^0 \in \mathbb{R}^n$, calculăm la fiecare iteraţie $k \geq 1$ un vector $x^k$ ale cărui componente sunt date de formula

$$x_i^k = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^k - \sum_{j=i+1}^{n} a_{ij} x_j^{k-1} \right) \qquad (1 \leq i \leq n), \tag{6.11}$$

sau, mai desfăşurat,

$$\begin{cases} x_1^k = \dfrac{1}{a_{11}} \left( b_1 - \displaystyle\sum_{j=2}^{n} a_{1j} x_j^{k-1} \right) \\ x_i^k = \dfrac{1}{a_{ii}} \left( b_i - \displaystyle\sum_{j=1}^{i-1} a_{ij} x_j^k - \sum_{j=i+1}^{n} a_{ij} x_j^{k-1} \right) \qquad (2 \leq i \leq n-1) \\ x_n^k = \dfrac{1}{a_{nn}} \left( b_n - \displaystyle\sum_{j=1}^{n-1} a_{nj} x_j^k \right). \end{cases} \tag{6.12}$$

**Observaţia 6.3.** *Pentru a putea defini şirul $(x^k)_{k\geq 0}$ prin (6.10) trebuie, ca şi la metoda Jacobi, să avem că $a_{ii} \neq 0$, pentru $i \in \{1, 2, .., n\}$. Această condiţie este echivalentă cu faptul că matricea $-E + D$ este inversabilă.*

**Observaţia 6.4.** *Spre deosebire de metoda Jacobi, unde componentele lui $x^k$ nu sunt luate in consideraţie decât la iteraţia $k+1$, la metoda Gauss-Seidel o componentă a lui $x^k$, odată calculată, este imediat folosită la calculul componentelor următoare ale lui $x^k$. Datorită acestei folosiri mai rapide a ultimelor calcule făcute, metoda lui Gauss-Seidel este, în multe situaţii dar nu întotdeauna, mai rapid convergentă decât metoda Jacobi.*

Avem următorul rezultat privind convergenţa metodei Gauss-Seidel.

**Teorema 6.2.** *Dacă $A$ este o matrice strict diagonal dominantă, atunci metoda iterativă Gauss-Seidel (6.10) este convergentă.*

*Proof.* Vom arăta că $\|B_G\|_\infty < 1$. Criteriul 2. de convergenţă a metodelor iterative enunţat la început ne va asigura că metoda Gauss-Seidel este convergentă. Fie

$$r = \max_{1\le i\le n}\left\{\frac{1}{|a_{ii}|}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\right\}.$$

Cum matricea $A$ este s.d.d., rezultă că $r < 1$. Vom demonstra că

$$\|B_G x\|_\infty \le r\|x\|_\infty \qquad (x \in \mathbb{C}^n). \tag{6.13}$$

Inegalitatea (6.13) implică

$$\|B_G\|_\infty \le r < 1,$$

şi prin urmare metoda Gauss-Seidel este convergentă.

Pentru a demonstra (6.13), fie $x \in \mathbb{C}^n$ şi notăm $y = B_G x$, ceea ce este echivalent cu

$$(-E + D)y = Fx. \tag{6.14}$$

Vom arăta prin inducţie după $i \in \{1, 2, ..., n\}$ că fiecare componentă $y_i$ a vectorului $y$ are proprietatea că $|y_i| \le r\|x\|_\infty$:

- Pentru $i = 1$, egalând prima componentă a vectorului din stânga cu prima componentă a vectorului din dreapta egalităţii (6.14), deducem că

  $$a_{11}y_1 = -\sum_{j=2}^{n} a_{1j}x_j.$$

  Prin urmare

  $$|y_1| \le \frac{1}{|a_{11}|}\left(\sum_{j=2}^{n}|a_{1j}|\,|x_j|\right) \le \frac{1}{|a_{11}|}\left(\sum_{j=2}^{n}|a_{1j}|\right)\|x\|_\infty \le r\|x\|_\infty,$$

  şi primul pas al inducţiei este verificat.

- Acum presupunem că $|y_j| \le r\|x\|_\infty$, $1 \le j \le i-1$, şi demonstrăm că $|y_i| \le r\|x\|_\infty$. Egalând componentele $i$ ale vectorilor din stânga şi din dreapta lui (6.14), deducem că

  $$\sum_{j=1}^{i} a_{ij}y_j = -\sum_{j=i+1}^{n} a_{ij}x_j,$$

  ceea ce este echivalent cu

  $$y_i = \frac{1}{a_{ii}}\left(-\sum_{j=1}^{i-1} a_{ij}y_j - \sum_{j=i+1}^{n} a_{ij}x_j\right).$$

  Prin urmare, folosind ipoteza de inducţie, obţinem că

  $$|y_i| \le \frac{1}{|a_{ii}|}\left(\sum_{j=1}^{i-1}|a_{ij}|\,|y_j| + \sum_{j=i+1}^{n}|a_{ij}||x_j|\right) \le \frac{1}{|a_{ii}|}\left(\sum_{j=1}^{i-1}|a_{ij}|\,r\|x\|_\infty + \sum_{j=i+1}^{n}|a_{ij}|\|x\|_\infty\right)$$

$$\leq \frac{1}{|a_{ii}|}\left(\sum_{j=1}^{i-1}|a_{ij}|+\sum_{j=i+1}^{n}|a_{ij}|\right)\|x\|_\infty \leq r\|x\|_\infty,$$

şi pasul de inducţie este verificat.

Am demonstrat deci că $|y_i| \leq r\|x\|_\infty$, pentru orice $i \in \{1, 2, \ldots, n\}$. Prin urmare,

$$\|B_G x\|_\infty = \|y\|_\infty = \max_{1\leq i\leq n}|y_i| \leq r\|x\|_\infty.$$

Am arătat că (6.13) se verifică şi demonstraţia se încheie. □

**Observaţia 6.5.** *Dacă matricea $A$ este s.d.d. atunci $a_{ii} \neq 0 > 0$, $1 \leq i \leq n$, ceea ce arată că matricea triunghiulară inferior $-E + D$ este inversabilă şi deci metoda Gauss-Seidel este bine definită.*

## 6.3 Exemple

**Exemplul 6.1.** *Folosind metoda Jacobi aproximaţi soluţia următorului sistem cu o eroare mai mică decât $10^{-1}$.*

$$\begin{cases} 10x_1 & -3x_2 & -x_3 & = 6 \\ -2x_1 & +10x_2 & +x_3 & = 9 \\ 2x_1 & -3x_2 & -10x_3 & = -11, \end{cases}$$

*folosind aproximaţia iniţială* $x^0 = \begin{pmatrix} x_1^0 \\ x_2^0 \\ x_3^0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

*Proof.* Avem că

$$A = \begin{pmatrix} 10 & -3 & -1 \\ -2 & 10 & 1 \\ 2 & -3 & -10 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 6 \\ 9 \\ -11 \end{pmatrix}.$$

Verificăm că matricea $A$ este s.d.d.:

$$\left.\begin{array}{l} |a_{11}| = 10 \\ |a_{12}| + |a_{13}| = |-3| + |-1| = 4 \end{array}\right\} \Rightarrow |a_{11}| > |a_{12}| + |a_{13}|$$

$$\left.\begin{array}{l} |a_{22}| = 10 \\ |a_{21}| + |a_{23}| = |-2| + |1| = 3 \end{array}\right\} \Rightarrow |a_{22}| > |a_{21}| + |a_{23}|$$

$$\left.\begin{array}{l} |a_{33}| = 10 \\ |a_{31}| + |a_{32}| = |2| + |-3| = 5 \end{array}\right\} \Rightarrow |a_{33}| > |a_{31}| + |a_{32}|.$$

Prin urmare $A$ este s.d.d şi metoda Jacobi pentru aproximarea soluţiei sistemului $Ax = b$ este convergentă. Să mai observăm şi că

$$\|B_J\|_\infty = \max\left\{\frac{2}{5}, \frac{3}{10}, \frac{1}{2}\right\} = \frac{1}{2}.$$

Voi arăta acum cum calculăm iteraţiile. Scriem sistemul în forma echivalentă:

$$\begin{cases} x_1 = \dfrac{1}{10}(3x_2 + x_3 + 6) \\ x_2 = \dfrac{1}{10}(2x_1 - x_3 + 9) \\ x_3 = -\dfrac{1}{10}(-2x_1 + 3x_2 - 11), \end{cases}$$

şi considerăm recurenţa

$$\begin{cases} x_1^k = \frac{1}{10}\left(3x_2^{k-1} + x_3^{k-1} + 6\right) \\ x_2^k = \frac{1}{10}\left(2x_1^{k-1} - x_3^{k-1} + 9\right) \\ x_3^k = -\frac{1}{10}\left(-2x_1^{k-1} + 3x_2^{k-1} - 11\right). \end{cases}$$

Observaţi că, am pus sus un indice $k$ în stânga egalului şi un indice $k-1$ în dreapta lui.

Pentru $k = 1$ obţinem

$$\begin{cases} x_1^1 = \frac{1}{10}\left(3x_2^0 + x_3^0 + 6\right) = \frac{1}{10}\left(0 + 0 + 6\right) = \frac{6}{10} \\ x_2^1 = \frac{1}{10}\left(2x_1^0 - x_3^0 + 9\right) = \frac{1}{10}\left(0 - 0 + 9\right) = \frac{9}{10} \\ x_3^1 = -\frac{1}{10}\left(-2x_1^0 + 3x_2^0 - 11\right) = -\frac{1}{10}\left(-0 + 0 - 11\right) = \frac{11}{10}. \end{cases}$$

Pentru $k = 2$ obţinem

$$\begin{cases} x_1^2 = \frac{1}{10}\left(3x_2^1 + x_3^1 + 6\right) = \frac{1}{10}\left(\frac{27}{10} + \frac{11}{10} + 6\right) = \frac{98}{100} \\ x_2^2 = \frac{1}{10}\left(2x_1^1 - x_3^1 + 9\right) = \frac{1}{10}\left(\frac{12}{10} - \frac{11}{10} + 9\right) = \frac{91}{100} \\ x_3^2 = -\frac{1}{10}\left(-2x_1^1 + 3x_2^1 - 11\right) = -\frac{1}{10}\left(-\frac{12}{10} + \frac{27}{10} - 11\right) = \frac{95}{100}. \end{cases}$$

Câţi termeni ai şirului trebuie să calculăm pentru a avea o aproximaţie cu o eroare mai mică decât $10^{-1}$? Fără să cunoaştem soluţia exactă, nu putem da un răspuns precis la această întrebare. Totuşi, putem obţine o estimare a numărului de iteraţii necesar folosind următoarea relaţie dată de Teorema 3.1.2 din cursul trecut:

$$\|x^* - x^k\|_\infty \leq \frac{\|B_J\|_\infty^k}{1 - \|B_J\|_\infty} \|x^1 - x^0\|_\infty.$$

Cum $\|B_J\|_\infty = \frac{1}{2}$ şi $\|x^1 - x^0\|_\infty = \max\{0.6, 0.9, 1.1\} = 1.1$ obţinem că

$$\|x^* - x^k\|_\infty \leq \frac{1.1}{2^{k-1}}.$$

Cum

$$\frac{1.1}{2^{k-1}} \leq 10^{-1} \Leftrightarrow 2^{k-1} \geq 11 \Leftrightarrow k \geq 5,$$

vom avea că,

$$\|x^* - x^5\|_\infty \leq 10^{-1}.$$

Deci, $x^5$ va reprezenta o aproximaţie cu o eroare mai mică decât $10^{-1}$ a soluţiei sistemului iniţial. Avem că

$$x^3 = \begin{pmatrix} 0.9680 \\ 1.0010 \\ 1.0230 \end{pmatrix}, \quad x^4 = \begin{pmatrix} 1.0026 \\ 0.9913 \\ 0.9933 \end{pmatrix}, \quad x^5 = \begin{pmatrix} 0.99672 \\ 1.00119 \\ 1.00313 \end{pmatrix}.$$

În final, să observăm că soluţia exactă este $x^* = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ iar $x^2$ reprezintă deja o aproximaţie cu o eroare de $0.05 < 10^{-1}$. $x^5$ este de fapt o aproximaţie mult mai bună, cu o eroare de 0.00328.

Acest lucru îl putem vedea aici pentru că am găsit soluţia exactă, lucru care în general nu este însă posibil.

□

**Exemplul 6.2.** *Aproximaţi soluţia sistemului de la Exemplul 6.1 cu o eroare mai mică decât* $10^{-1}$, *folosind metoda Gauss-Seidel şi aproximaţia iniţială* $x^0 = \begin{pmatrix} x_1^0 \\ x_2^0 \\ x_3^0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

*Proof.* Ştim deja că matricea $A$ este s.d.d. şi metoda Gauss-Seidel pentru aproximarea soluţiei sistemului $Ax = b$ este convergentă.

Voi arăta acum cum calculăm iteraţiile. Scriem sistemul în forma echivalentă:

$$\begin{cases} x_1 = \dfrac{1}{10}(3x_2 + x_3 + 6) \\ x_2 = \dfrac{1}{10}(2x_1 - x_3 + 9) \\ x_3 = -\dfrac{1}{10}(-2x_1 + 3x_2 - 11), \end{cases}$$

şi considerăm recurenţa

$$\begin{cases} x_1^k = \dfrac{1}{10}\left(3x_2^{k-1} + x_3^{k-1} + 6\right) \\ x_2^k = \dfrac{1}{10}\left(2x_1^k - x_3^{k-1} + 9\right) \\ x_3^k = -\dfrac{1}{10}\left(-2x_1^k + 3x_2^k - 11\right). \end{cases}$$

Observaţi că, am pus sus indici $k$ în stânga egalului dar şi în dreapta lui, la elementele care au jos un indice mai mic decât numărul ecuaţiei.

Pentru $k = 1$ obţinem

$$\begin{cases} x_1^1 = \dfrac{1}{10}\left(3x_2^0 + x_3^0 + 6\right) = \dfrac{1}{10}\left(0 + 0 + 6\right) = \dfrac{6}{10} \\ x_2^1 = \dfrac{1}{10}\left(2x_1^1 - x_3^0 + 9\right) = \dfrac{1}{10}\left(\dfrac{12}{10} - 0 + 9\right) = \dfrac{102}{100} \\ x_3^1 = -\dfrac{1}{10}\left(-2x_1^1 + 3x_2^1 - 11\right) = -\dfrac{1}{10}\left(-\dfrac{12}{10} + \dfrac{306}{100} - 11\right) = \dfrac{914}{1000}. \end{cases}$$

Pentru $k = 2$ obţinem

$$\begin{cases} x_1^2 = \dfrac{1}{10}\left(3x_2^1 + x_3^1 + 6\right) = \dfrac{1}{10}\left(\dfrac{306}{100} + \dfrac{914}{1000} + 6\right) = \dfrac{9974}{10000} \\ x_2^2 = \dfrac{1}{10}\left(2x_1^2 - x_3^1 + 9\right) = \dfrac{1}{10}\left(\dfrac{19948}{10000} - \dfrac{914}{1000} + 9\right) = \dfrac{100808}{100000} \\ x_3^2 = -\dfrac{1}{10}\left(-2x_1^2 + 3x_2^2 - 11\right) = -\dfrac{1}{10}\left(-\dfrac{19948}{10000} + \dfrac{302424}{100000} - 11\right) = \dfrac{997056}{1000000}. \end{cases}$$

Cât timp vom calcula? Pentru a da o estimare a numărului de iteraţii necesare folosim relaţia dată de Teorema 3.1.2 din cursul trecut:

$$\|x^* - x^k\|_\infty \leq \frac{\|B_G\|_\infty^k}{1 - \|B_G\|_\infty} \|x^1 - x^0\|_\infty.$$

Determinăm matricea $B_G$:

$$B_G = \begin{pmatrix} 10 & 0 & 0 \\ -2 & 10 & 0 \\ 2 & -3 & -10 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -3 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ \frac{2}{100} & \frac{1}{10} & 0 \\ \frac{14}{1000} & -\frac{3}{100} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} 0 & -3 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -\frac{3}{10} & -\frac{1}{10} \\ 0 & -\frac{6}{100} & \frac{8}{100} \\ 0 & -\frac{42}{1000} & -\frac{44}{1000} \end{pmatrix}.$$

Avem că $\|B_G\|_\infty = \max\left\{\frac{4}{10}, \frac{14}{100}, \frac{86}{1000}\right\} = \frac{2}{5}$ şi $\|x^1 - x^0\|_\infty = \max\{0.6, 1.02, 0.914\} = 1.02$ şi obţinem că următoarea estimare are loc

$$\|x^* - x^k\|_\infty \leq \frac{5.1}{3}\left(\frac{2}{5}\right)^k.$$

Cum

$$\frac{5.1}{3}\left(\frac{2}{5}\right)^k \leq 10^{-1} \Leftrightarrow \left(\frac{5}{2}\right)^k \geq 17 \Leftrightarrow k \geq 4,$$

vom avea că,

$$\|x^* - x^4\|_\infty \leq 10^{-1}.$$

Deci, am arătat că $x^4$ va reprezenta o aproximaţie cu o eroare mai mică decât $10^{-1}$ a soluţiei sistemului iniţial. Avem că

$$x^3 = \begin{pmatrix} 1.0021296 \\ 1.00072032 \\ 1.000209824 \end{pmatrix}, \quad x^4 = \begin{pmatrix} 1.0002370784 \\ 1.00002643328 \\ 1.000039485696 \end{pmatrix}.$$

Cum soluţia exactă este $x^* = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, vedem că $x^2$ reprezintă deja o aproximare cu o eroare de 0.00808 iar $x^4$ este o aproximaţie cu o eroare mai mică decât 0.00024, dar acest lucru nu-l puteam şti dacă nu aveam soluţia $x^*$. □

# Cursul 7

# Probleme de optimizare

## 7.1 Introducere

Fie o mulţime $\Omega \subseteq \mathbb{R}^N$ şi o funcţie $f : \Omega \to \mathbb{R}$ definită în $\Omega$ şi cu valori reale.

**Definiţia 7.1.** *$x^* \in \Omega$ este un* **punct de minim (maxim) global** *al funcţiei $f$ dacă*

$$f(x^*) \leq f(x), \quad \forall x \in \Omega, \qquad (\textit{respectiv } \geq). \tag{7.1}$$

Problema găsirii unui punct de minim $x^*$ al funcţiei $f$ în $\Omega$ se scrie prescurtat astfel

$$f(x^*) = \min_{x \in \Omega} f(x). \tag{7.2}$$

**Definiţia 7.2.** *$x^* \in \Omega$ este un* **punct de minim (maxim) local** *al funcţiei $f$ dacă există o vecinătate $V$ a lui $x^*$ astfel ca*

$$f(x^*) \leq f(x), \quad \forall x \in V \cap \Omega, \qquad (\textit{respectiv } \geq). \tag{7.3}$$

Un punt de minim sau de maxim este un **punct de extrem** al funcţiei $f$. Se mai foloseşte terminologia de punct de extrem relativ (în loc de local) şi de punct de extrem absolut (în loc de global).

**Observaţia 7.1.** *O problemă de maxim se poate reduce întotdeauna la una de minim datorită relaţiei:*

$$\max_{x \in \Omega} f(x) = \min_{x \in \Omega} (-f(x)).$$

*De aceea, ne vom opri doar asupra studiului problemelor de minim.*

Problema găsirii minimelor unei funcţii este o problemă de optimizare. Într-o astfel de problemă apar câteva întrebări:

- Există un minim pentru $f$?
- Avem minim global sau local?
- Este minimul unic?
- Cum putem găsi sau aproxima minimul?

### 7.1.1 Funcţii convexe

Funcţiile convexe au o mare importanţă în problemele de optimizări deoarece apar relativ frecvent şi au o serie de proprietăţi specifice care uşurează analiza problemei. Reamintim următoarele două definiţii.

**Definiţia 7.3.** *O mulţime $\Omega \subseteq \mathbb{R}^N$ este* **convexă** *dacă, pentru orice $u, v \in \Omega$ şi $\lambda \in [0,1]$, avem că*

$$\lambda u + (1-\lambda)v \in \Omega. \tag{7.4}$$

**Definiţia 7.4.** *Fie mulţimea convexă $\Omega \subseteq \mathbb{R}^N$. O funcţie $f : \Omega \to \mathbb{R}$ se numeşte* **convexă** *dacă, pentru orice $u, v \in \Omega$ şi $\lambda \in [0,1]$, avem că*

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v). \tag{7.5}$$

*Dacă inegalitatea de mai sus este strică pentru orice $\lambda \in (0,1)$ şi $u \neq v$, atunci $f$ se numeşte* **strict convexă**.

Următoarele proprietăţi importante ale funcţiilor convexe sunt cunoscute de la cursul de Cercetări operaţionale.

**Teorema 7.1.** *Fie $f : \Omega \to \mathbb{R}$ o funcţie convexă pe mulţimea convexă $\Omega \subseteq \mathbb{R}^N$. Atunci*

1. *Dacă $\Omega$ este deschisă atunci $f$ este continuă în $\Omega$.*
2. *Orice minim local este minim global al lui $f$.*
3. *Dacă în plus $f$ este strict convexă, există cel mult un minim global.*

*Proof.* i) Vezi [N].

ii) Fie $x \in \Omega$ un minim local al lui $f$ şi $V$ vecinătate a lui $x$ astfel ca

$$f(x) \leq f(t), \quad \forall t \in V \cap \Omega.$$

Presupunem că există $y \in \Omega$ astfel ca $f(y) < f(x)$. Pentru orice $\lambda \in (0,1)$ avem că

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x) < f(x).$$

Dar $x + \lambda(y-x) = \lambda y + (1-\lambda)x \in \Omega$ şi pentru $\lambda$ suficient de mic $x + \lambda(y-x) \in V$ ceea ce conduce la o contradicţie.

iii) Fie $x, y \in \Omega$ două minime ale lui $f$. Dacă $x \neq y$, luând $\lambda \in (0,1)$, avem că

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) = \min_{t \in \Omega} f(t)$$

ceea ce reprezintă o contradicţie. □

Importantă este şi următoarea caracterizare a funcţiilor convexe.

**Teorema 7.2.** *Fie $f$ o funcţie diferenţiabilă definită pe mulţimea convexă $\Omega \subseteq \mathbb{R}^N$ şi cu valori în $\mathbb{R}$. Următoarele afirmaţii sunt echivalente:*

*i.) $f$ este convexă.*

*ii.) $f$ verifică inegalitatea*

$$f(v) - f(w) \geq (\nabla f(w), v - w), \quad \forall v, w \in \Omega. \tag{7.6}$$

*Proof.* "$i) \Rightarrow ii)$"

Din ipoteza de convexitate avem că

$$f\left((1-\lambda)\,w+\lambda v\right)\leq(1-\lambda)\,f(w)+\lambda f(v),$$

de unde rezultă că

$$\frac{1}{\lambda}\left[f\left((1-\lambda)\,w+\lambda v\right)-f(w)\right]\leq f(v)-f(w).$$

Facând $\lambda\to 0$ în relaţia de mai sus şi ţinând cont că

$$\lim_{h\to 0}\frac{f(w+h(v-w))-f(w)}{h}=(\nabla f(w),v-w),$$

obţinem (7.6).

"$ii) \Rightarrow i)$" Fie $u,z\in\Omega$ şi $\lambda\in[0,1]$.

Scriem relaţia (7.6) cu $v=z$ şi $w=\lambda u+(1-\lambda)z\in\Omega$:

$$f(z)\geq f(\lambda u+(1-\lambda)z)-\lambda(\nabla f(\lambda u+(1-\lambda)z),u-z).$$

Apoi, scriem relaţia (7.6) cu $v=u$ şi $w=\lambda u+(1-\lambda)z$:

$$f(u)\geq f(\lambda u+(1-\lambda)z)+(1-\lambda)(\nabla f(\lambda u+(1-\lambda)z),u-z).$$

Din cele două relaţii obţinute, avem că

$$\lambda f(u)+(1-\lambda)f(z)\geq f(\lambda u+(1-\lambda)z),$$

ceea ce ne arată că $f$ este convexă. □

**Observaţia 7.2.** *Semnificaţia geometrică a relaţiei* (7.6) *este aceea că, pentru o funcţie convexă, graficul se află deasupra planului tangent. În dimensiune unu,* (7.6) *se poate scrie*

$$f(x)\geq f(x_0)+f'(x_0)(x-x_0),$$

*iar* $y=f(x_0)+f'(x_0)(x-x_0)$ *reprezintă exact ecuaţia dreptei tangente la graficul lui* $f$ *în punctul* $(x_0,f(x_0))$.

### 7.1.2 Funcţionale eliptice

Vom considera $J:\Omega\subset\mathbb{R}^N\to\mathbb{R}$.

**Definiţia 7.5.** *Funcţia* $J$ *se numeşte* **coercivă** *dacă*

$$\lim_{\substack{\|v\|\to\infty\\ v\in\Omega}}J(v)=\infty. \tag{7.7}$$

Observaţi că definiţia de mai sus are sens doar dacă $\Omega$ este o mulţime nemărginită.

**Definiţia 7.6.** *Funcţia* $J$ *se numeşte* **eliptică** *(sau* $\alpha$**-eliptică**$)$ *dacă este de diferenţiabilă în* $\Omega$ *şi există* $\alpha>0$ *astfel încât*

$$(\nabla J(v)-\nabla J(w),v-w)\geq\alpha||v-w||^2,\quad\forall v,w\in\Omega. \tag{7.8}$$

Un exemplu important de funcţională eliptică este următorul.

**Propoziţia 7.1.** *Fie $A \in \mathcal{M}_N(\mathbb{R})$ o matrice simetrică şi definită pozitiv şi $b \in \mathbb{R}^N$. Funcţionala $J : \mathbb{R}^N \to \mathbb{R}$, definită prin*

$$J(v) = \frac{1}{2}(Av, v) - (b, v), \tag{7.9}$$

*este eliptică.*

*Proof.* Evident, $J$ este de clasă $C^1(\mathbb{R}^N)$ şi $\nabla J(v) = Av - b$. Rezultă că

$$(\nabla Jv - \nabla Jw, v - w) = (A(v - w), v - w) =$$

$$= \frac{(A(v-w), v-w)}{(v-w, v-w)} ||v - w||^2 \geq \min_{x \in \mathbb{C}^N} \frac{(Ax, x)}{(x, x)} ||v - w||^2 = \lambda_1 ||v - w||^2,$$

unde $\lambda_1 > 0$ este cea mai mică valoare proprie a lui $A$. Pentru ultima egalitate am folosit Teorema Courant-Fischer de caracterizare a valorilor proprii ale unei matrici hermitice.

Deci, pentru orice $v, w \in \mathbb{R}^N$, avem că

$$(\nabla Jv - \nabla Jw, v - w) \geq \lambda_1 ||v - w||^2,$$

şi $J$ este eliptică. □

O noţiune mai puternică decât convexitatea este următoarea.

**Definiţia 7.7.** *Fie mulţimea convexă $\Omega \subseteq \mathbb{R}^N$. O funcţie $f : \Omega \to \mathbb{R}$ se numeşte* **uniform convexă** *(sau $\alpha-$***convexă***) dacă există $\alpha > 0$ astfel încât, pentru orice $\lambda \in [0, 1]$ şi $u, v \in \Omega$, avem că*

$$f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda) f(v) - \frac{\alpha}{2}\lambda(1 - \lambda)||u - v||^2. \tag{7.10}$$

Să observăm că o funcţie uniform convexă este strict convexă. Următorul rezultat caracterizează funcţiile uniform convexe.

**Teorema 7.3.** *Fie $J$ o funcţie diferenţiabilă definită pe mulţimea convexă $\Omega \subseteq \mathbb{R}^N$ şi cu valori în $\mathbb{R}$. Următoarele afirmaţii sunt echivalente:*

1. *$J$ este $\alpha-$convexă.*
2. *$J$ verifică inegalitatea*

$$J(v) - J(w) \geq (\nabla J(w), v - w) + \frac{\alpha}{2}||v - w||^2, \quad \forall v, w \in \Omega. \tag{7.11}$$

3. *$J$ este $\alpha-$eliptică:*

$$(\nabla J(v) - \nabla J(w), v - w) \geq \alpha||v - w||^2, \quad \forall v, w \in \Omega.$$

*Proof.* "$i) \Rightarrow ii)$" Din ipoteza de $\alpha-$convexitate avem că

$$J((1 - \lambda) w + \lambda v) \leq (1 - \lambda) J(w) + \lambda J(v) - \frac{\alpha}{2}\lambda(1 - \lambda)||w - v||^2.$$

de unde rezultă că

$$\frac{1}{\lambda}[J((1 - \lambda) w + \lambda v) - J(w)] \leq J(v) - J(w) - \frac{\alpha}{2}(1 - \lambda)||w - v||^2.$$

Facând $\lambda \to 0$ în relaţia de mai sus şi ţinând cont că

$$\lim_{h\to 0} \frac{J(w+h(v-w))-J(w)}{h} = (\nabla J(w), v-w),$$

obţinem (7.11).

"*ii*) $\Rightarrow$ *iii*)" Scriem relaţia (7.11) schimbând locul lui $v$ cu $w$ şi adunăm la (7.11). Se obţine imediat (7.8).

"*iii*) $\Rightarrow$ *i*)" Date $u, v \in \Omega$, definim funcţia $\varphi : [0,1] \to \mathbb{R}$ prin $\varphi(t) = J(u+t(v-u))$. Funcţia $\varphi$ este derivabilă şi

$$\varphi'(t) = (\nabla J(u+t(v-u)), v-u).$$

Prin urmare, pentru $t > s$,

$$\varphi'(t) - \varphi'(s) = (\nabla J(u+t(v-u)) - \nabla J(u+s(v-u)), v-u)) =$$

$$= \frac{1}{t-s} \left(\nabla J(u+t(v-u)) - \nabla J(u+s(v-u)), (u+t(v-u)) - (u+s(v-u))\right).$$

Folosind ipoteza de elipticitate a lui $J$ obţinem că

$$\varphi'(t) - \varphi'(s) \geq \alpha(t-s)||v-u||^2 \quad \forall t, s \in [0,1] \;\; t \geq s.$$

Integrând ultima inegalitate în $t \in (\lambda, 1)$ şi în $s \in (0, \lambda)$, obţinem că

$$\lambda\varphi(1) + (1-\lambda)\varphi(0) - \varphi(\lambda) \geq \frac{\alpha}{2}\lambda(1-\lambda)||v-u||^2$$

şi prin urmare

$$\lambda J(v) + (1-\lambda)J(u) - J(\lambda v + (1-\lambda)u) \geq \frac{\alpha}{2}\lambda(1-\lambda)||v-u||^2.$$

Am demonstrat astfel că $J$ este $\alpha-$convexă.

"*iii*) $\Rightarrow$ *ii*)" (poate fi omisă- este o variantă care nu foloseşte $\alpha-$convexitatea.)

Aplicând formula lui Taylor cu restul în formă integrală, rezultă că

$$J(v) = J(w) + \int_0^1 (\nabla J(w+t(v-w)), v-w)dt =$$

$$= J(w) + (\nabla J(w), v-w) + \int_0^1 \left[(\nabla J(w+t(v-w)), v-w) - (\nabla J(w), v-w)\right] dt \geq$$

$$J(w) + (\nabla J(w), v-w) + \int_0^1 \alpha t||v-w||^2 dt = J(w) + (\nabla J(w), v-w) + \frac{\alpha}{2}||v-w||^2$$

şi (7.11) este demonstrată.

□

### 7.1.3 Criterii de găsire a punctelor de minim

Vom începe prin prezentarea a două teoreme importante care asigură existenţa minimului.

**Teorema 7.4.** *(Weierstrass) Fie $U \subseteq \mathbb{R}^N$ o mulţime compactă şi nevidă şi $J : U \to \mathbb{R}$ o funţională continuă. Atunci există $u \in U$ astfel ca*

$$J(u) = \min_{v \in U} J(v). \tag{7.12}$$

Reamintim că o mulţime $U \subseteq \mathbb{R}^N$ este compactă dacă şi numai dacă este mărginită şi închisă. În teorema de mai sus ipoteza de mărginire a lui $U$ este esenţială. Totuşi, ea poate fi eliminată dacă funcţionala $J$ are o proprietate adiţională.

**Teorema 7.5.** *Fie $U \subseteq \mathbb{R}^N$ o mulţime închisă şi nevidă şi $J : U \to \mathbb{R}$ o funcţională continuă şi coercivă dacă $U$ este nemărginită. Atunci există $u \in U$ astfel ca*

$$J(u) = \min_{v \in U} J(v). \tag{7.13}$$

*Proof.* Dacă $U$ mărginită, atunci ea este compactă şi aplicăm Teorema Weierstrass. Dacă $U$ este nemărginită, fie $v^0 \in U$ şi definim

$$A = \{x \in U \, : \, J(x) \leq J(v^0)\}.$$

$A$ este mărginită (din proprietatea de coercivitate a lui $J$) şi închisă deoarece

$$A = U \cap J^{-1}((-\infty, J(v^0)]).$$

Prin urmare, $A$ este compactă şi $J$ îşi atinge minimul în $A$. Fie $J(u) = \min_{v \in A} J(v)$. Cum, pentru orice $v \in U \setminus A$, $J(v) > J(v^0) \geq J(u)$ rezultă că de fapt $u$ este minimul lui $J$ în $U$. □

Următorul rezultat este cunoscut şi foarte important deoarece reduce problema găsirii minimului la determinarea punctelor critice.

**Teorema 7.6.** *Fie $U \subseteq \mathbb{R}^N$ o mulţime deschisă şi $f : U \to \mathbb{R}$ o funcţie continuă pe $U$ şi derivabilă în $u^*$. Dacă $u^*$ este un punct de minim local, atunci*

$$\nabla f(u^*) = 0. \tag{7.14}$$

*Dacă $f$ este convexă, atunci condiţia (7.14) implică faptul că $u^*$ este punct de minim al lui $f$ în $U$.*

**Observaţia 7.3.** *Teorema 7.6 ne arată că punctele de minim sunt printre punctele critice ale funcţiei. Condiţia ca mulţimea $U$ să fie deschis este esenţială în Teorema 7.6 (de exemplu, funcţia $f(x) = x$ pe $[0, 1]$ are un minim în $x^* = 0$ dar derivata nu se anulează).*

Ca o aplicaţie a Teoremei 7.6 să notăm următorul rezultat.

**Propoziţia 7.2.** *Fie $A \in \mathcal{M}_N(\mathbb{R})$ o matrice simetrică şi definită pozitiv şi $b \in \mathbb{R}^N$. Sistemul*

$$Ax = b, \tag{7.15}$$

*are o unică soluţie care coincide cu punctul de minim al funcţionalei $J : \mathbb{R}^N \to \mathbb{R}$, definită prin*

$$J(v) = \frac{1}{2}(Av, v) - (b, v). \tag{7.16}$$

.

*Proof.* Din Propoziţia 7.1 ştim că $J$ este eliptică în $\mathbb{R}^N$ şi $\nabla J(v) = Av - b$. Conform Teoremei 7.3, avem că $J$ este uniform convexă. Prin urmare există un unic punct de minim al lui $J$ care verifică sistemul (7.2). □

Pentru mulţimi convexe, avem un rezultat care este valabil şi pe mulţimi care nu sunt deschise, aşa cum cere Teorema 7.6. El generalizează următoarea observaţie elementară: dacă $u \in [a, b]$ este punct de minim al funcţiei derivabile $f$ în $[a, b] \subset \mathbb{R}$ atunci

$$\begin{cases} f'(u) \geq 0 & \text{dacă } u = a \\ f'(u) = 0 & \text{dacă } u \in (a, b) \\ f'(u) \leq 0 & \text{dacă } u = b \end{cases} \Leftrightarrow f'(u)(x - u) \geq 0 \quad \forall x \in [a, b].$$

Mai mult, dacă $f$ este convexă, ultima inegalitate este o condiţie suficientă pentru minim.

**Teorema 7.7.** *Fie date $U \subset \mathbb{R}^N$ o mulţime nevidă şi convexă, $f : U \to \mathbb{R}$ şi $u \in U$ un punct în care $f$ este derivabilă. Dacă $u$ este un punct de minim global pentru $f$ în $U$, atunci avem că*

$$(\nabla f(u), (v - u)) \geq 0, \ \forall v \in U. \tag{7.17}$$

*Reciproc, dacă $u \in U$ verifică (7.17) şi $f$ este convexă atunci $u$ este un punct de minim global pentru $f$ în $U$.*

*Proof.* Cum $U$ este convexă, pentru orice $\lambda \in (0, 1)$ şi $v \in U$, avem că $u + \lambda(v - u) \in U$. Cum $u$ este punct de minim al lui $f$, obţinem că

$$f(u) \leq f(u + \lambda(v - u)).$$

Rezultă că, pentru orice $v \in U$,

$$0 \leq \lim_{\lambda \to 0} \frac{1}{\lambda} \left( f(u + \lambda(v - u)) - f(u) \right) = (\nabla f(u), (v - u)),$$

şi (7.17) este demonstrată,

Reciproc, dacă $u \in U$ verifică (7.17) şi $f$ este convexă, din Teorema 7.2 rezultă că

$$f(v) - f(u) \geq (\nabla f(u), (v - u)) \geq 0,$$

pentru orice $v \in U$ şi deci $u$ este punct de minim global al lui $f$. □

**Observaţia 7.4.** *Dacă $U$ este deschisă atunci*

$$(\nabla f(u), (v - u)) \geq 0, \ \forall v \in U \Leftrightarrow \nabla f(u) = 0.$$

*Într-adevăr, cum $U$ este deschisă, există o bilă $B(u, r) \subset U$. Pentru orice $w \in \mathbb{R}^N \setminus \{0\}$, avem că $v = u + \frac{r}{2} \frac{w}{||w||} \in B(u, r)$. Prin urmare, pentru orice $w \in \mathbb{R}^N$,*

$$(\nabla f(u), w) = \frac{2||w||}{r} (\nabla f(u), v - u) = 0$$

*şi deci $\nabla f(u) = 0$.*

Avem următoarea proprietate importantă a unei funcţii eliptice.

**Teorema 7.8.** *Dacă $U \subseteq \mathbb{R}^N$ este nevidă, închisă şi convexă iar $J$ este eliptică, atunci există un unic $u \in U$ astfel ca*

$$J(u) = \min_{v \in U} J(v). \tag{7.18}$$

*În plus, elementul $u$ este caracterizat de inegalitatea*

$$(\nabla J(u), v - u) \geq 0, \quad \forall v \in U. \tag{7.19}$$

*Proof.* Luând $w = 0$ în (7.11), rezultă că

$$J(v) \geq J(0) + (\nabla J(0), v) + \frac{\alpha}{2}||v||^2 \geq J(0) - ||\nabla J(0)||\,||v|| + \frac{\alpha}{2}||v||^2$$

de unde rezultă coercivitatea lui $J$. Existenţa punctului de minim rezultă din Teorema 7.5.

Conform Teoremei 7.3, fiind eliptică $J$ este şi uniform convexă. Prin urmare $J$ este strict convexă, de unde rezultă unicitatea punctului de minim. De asemenea, caracterizarea (7.19) a acestui punct de minim este cea întâlnită deja în Teorema 7.7. □

# Cursul 9

# Probleme de optimizare Metode de coborâre. Metoda gradientului

Existenţa şi caracterizarea punctelor de minim ale funcţionalelor cu valori reale au fost prezentate la începutul acestui capitol. Începem acum să introducem principalele metode de aproximare a lor. Vom presupune în această secţiune că $J : \mathbb{R}^N \to \mathbb{R}$ este o funcţională eliptică. Conform rezultatelor din Cursul 7, există un unic punct de minim $x^* \in \mathbb{R}^N$ al lui $J$:

$$J(x^*) = \min_{x \in \mathbb{R}^N} J(x). \tag{9.1}$$

În plus ştim că punctul de minim $x^*$ este caracterizat de relaţia

$$\nabla J(x^*) = 0. \tag{9.2}$$

## 9.1 Metode de coborâre

Metodele de coborâre au la bază următoarea strategie pentru aproximarea lui $x^*$:

- Se alege $x^0 \in \mathbb{R}^N$;
- Se construieşte recursiv şirul $(x^k)_{k \geq 0}$

$$x^{k+1} = x^k + r_k d^k \quad (k \geq 0), \tag{9.3}$$

unde $r_k \in \mathbb{R}$ se numeşte pas de coborâre iar $d^k \in \mathbb{R}^N$ direcţie de coborâre.

Pasul de coborâre se poate lua constant, se poate modifica arbitrar sau se poate optimiza la fiecare iteraţie $k$. În primul caz avem de aface cu o metodă de **pas fix** în cel de-al doilea caz cu o metodă de **pas variabil** iar în ultimul caz cu o metodă de **pas optimal**.

## 9.2 Determinarea pasului optimal pentru o direcţie $d$

Pasul optimal este acea valoare $r^*$ cu proprietatea că

$$J(x+r^*d) = \min_{r\in\mathbb{R}} J(x+rd). \tag{9.4}$$

Să observăm că prin (9.4) se minimizează funcţia de o variabilă reală $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(r) = J(x+rd)$. Privind dintr-un alt unghi, avem o minimizare a lui $J$ pe submulţimea închisă, convexă şi nevidă $U_d = \{x + rd : r \in \mathbb{R}\}$, pe care o putem interpreta ca o dreaptă ce trece prin $x$ şi are direcţia $d$. Dacă $J$ este eliptică în $\mathbb{R}^N$, atunci este asigurată existenţa şi unicitatea lui $r^*$. Cum de fapt avem o problemă de minimizare a funcţiei $\varphi(r)$ definită pe dreapta reală, minimul va fi un punct critic al lui $\varphi$, ceea ce este echivalent cu

$$0 = \varphi'(r^*) = \frac{d}{dr} J(x+rd)\Big|_{r=r^*} = (\nabla J(x+r^*d), d). \tag{9.5}$$

Ecuaţia (9.5) poate să determine o formulă explicită a lui $r^*$. Să vedem, de exemplu, care este formula lui $r^*$ în cazul particular

$$J(x) = \frac{1}{2}(Ax, x) - (b, x).$$

Cum $\nabla J(x) = Ax - b$, din (9.5) rezultă că $r^*$ verifică $(A(x + r^*d) - b, d) = 0$, şi prin urmare

$$r^* = \frac{(-Ax+b, d)}{(Ad, d)} = -\frac{(\nabla J(x), d)}{(Ad, d)}. \tag{9.6}$$

În cazul unei funcţii $J$ mai generale, pentru găsirea lui $r^*$ se pot folosi metode de minimizare a funcţiilor de o variabilă reală, precum metoda de căutare pe baza secţiunii de aur.

Alegerea direcţiei de coborâre poate fi făcută în mai multe feluri, în funcţie de acesta având o metodă sau alta.

## 9.3 Metoda gradientului

Aceasta este o metodă de coborâre în care direcţia de coborâre se alege după cum urmează:

$$d^k = -\nabla J(x^k), \quad k \geq 0. \tag{9.7}$$

Prin urmare, în metoda gradientului direcţia de căutare $d^k$ este calculată la fiecare pas în funcţie de $x^k$ şi reprezintă geometric **direcţia în care funcţia $J$ descreşte cel mai repede**. Pasul de coborâre poate fi luat fix sau variabil. Următoarea teoremă ne arată că metoda gradientului cu pas optimal este convergentă dacă $J$ este eliptică.

**Teorema 9.1.** *(Metoda gradientului cu pas optimal) Fie $J : \mathbb{R}^N \to \mathbb{R}$ o funcţie de clasă $C^1(\mathbb{R}^N)$ eliptică, adică o funcţie cu proprietatea că există $\alpha > 0$ astfel încât:*

$$(\nabla J(v) - \nabla J(w), v - w) \geq \alpha\|v - w\|^2, \ \forall v, w \in \mathbb{R}^N. \tag{9.8}$$

*Atunci şirul dat de (9.3) cu pas optimal (9.4) şi direcţia (9.7) converge la $x^*$, punctul în care $J$ îşi atinge valoarea minimă.*

*Proof.* Din ipoteza de elipticitate a lui $J$ ştim că există un unic punct de minim $x^* \in \mathbb{R}^N$ al lui $J$. Dacă $\nabla J(x^k) = 0$ am găsit minumul care este chiar $x^k$. Putem presupune că $\nabla J(x^k) \neq 0$ pentru orice $k \geq 0$. Funcţia

$$\varphi : r \to \varphi(r) = J(x^k - r\nabla J(x^k))$$

este coercivă, continuă şi strict convexă, prin urmare are un singur minim $r_k^*$ cu $\varphi'(r_k^*) = 0$. Acesta este pasul optimal şi avem că $x^{k+1} = x^k - r_k^*\nabla J(x^k)$. Cum $r_k^*$ este minimul lui $\varphi$ în $\mathbb{R}$, rezultă că

$$0 = \varphi'(r_k^*) = -(\nabla J(x^k - r_k^*\nabla J(x^k)), \nabla J(x^k))$$

şi prin urmare

$$(\nabla J(x^{k+1}), \nabla J(x^k)) = 0 \tag{9.9}$$

ceea ce demonstrază că două direcţii consecutive sunt ortogonale.

De asemenea, din $x^{k+1} = x^k - r_k^*\nabla J(x^k)$ deducem că

$$(\nabla J(x^{k+1}), x^{k+1} - x^k) = 0. \tag{9.10}$$

Ne reamintim că, pentru o funcţie $\alpha-$eliptică, are loc inegalitatea:

$$J(v) - J(w) \geq (\nabla J(w), v - w) + \frac{\alpha}{2}\|v - w\|^2, \quad \forall v, w \in \Omega. \tag{9.11}$$

Prin urmare din (9.10) şi (9.11) rezultă că

$$\frac{\alpha}{2}\|x^{k+1} - x^k\|^2 \leq J(x^k) - J(x^{k+1}). \tag{9.12}$$

Rezultă că şirul $(J(x^k))_{k\geq 0}$ este descrescător şi, fiind şi minorat de $J(x^*)$, deducem că

$$\lim_{k\to\infty} [J(x^{k+1}) - J(x^k)] = 0$$

şi prin urmare

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{9.13}$$

Cum $J$ este coercivă iar şirul $(J(x^k))_k$ este descrescător rezultă că $(x^k)_k$ este mărginit. Derivata $\nabla J$ fiind continuă este uniform continuă pe compacte şi din (9.13) rezultă că

$$\lim_{k\to\infty} \|\nabla J(x^{k+1}) - \nabla J(x^k)\| = 0. \tag{9.14}$$

Deoarece

$$\|\nabla J(x^k)\|^2 = (\nabla J(x^k), \nabla J(x^k)) = (\nabla J(x^k), \nabla J(x^k) - \nabla J(x^{k+1})) \leq \|J(x^k)\| \, \|\nabla J(x^k) - \nabla J(x^{k+1})\|,$$

din (9.14) deducem că

$$\lim_{k\to\infty} \nabla J(x^k) = 0. \tag{9.15}$$

Acum putem demonstra convergenţa. Folosind ipoteza de elipticitate a lui $J$ şi faptul că $\nabla J(x^*) = 0$, se obţine că

$$\alpha\|x^k - x^*\|^2 \leq (\nabla J(x^k) - \nabla J(x^*), x^k - x^*) = (\nabla J(x^k), x^k - x^*) \leq \|\nabla J(x^k)\| \, \|x^k - x^*\|$$

şi prin urmare

$$\|x^k - x^*\| \leq \frac{1}{\alpha}\|\nabla J(x^k)\|. \tag{9.16}$$

Convergenţa şirului $(x^k)_{k\geq 0}$ la $x^*$ rezultă acum din (9.15). □

**Observaţia 9.1.** *Inegalitatea (9.16) oferă un criteriu de oprire a iteraţiilor: dat $\varepsilon > 0$, calculăm iteraţii $x^k$ până când $\|\nabla J(x^k)\| < \varepsilon$.*

Următorul caz particular este foarte important:

$$J(v) = \frac{1}{2}(Av, v) - (b, v) + c \tag{9.17}$$

unde $A \in \mathcal{M}_N(\mathbb{R})$ este o matrice simetrică definită pozitiv, $b \in \mathbb{R}^N$ şi $c \in \mathbb{R}$. Amintim că, în acest caz, $J$ este $\alpha-$eliptică pe $\mathbb{R}^N$, cu $\alpha = \lambda_1 = \min\{\lambda : \lambda \in \sigma(A)\}$. În acest caz particular, o iteraţie a metodei gradientului cu pas optimal se calculează astfel:

- $g^k = Ax^k - b$
- $r_k = \dfrac{\|g^k\|^2}{(Ag^k, g^k)}$
- $x^{k+1} = x^k - r_k g^k$

Următorul rezultat arată convergenţa metodei gradientului cu pas fix sau variabil în anumite ipoteze suplimentare.

**Teorema 9.2.** *Fie $J : \mathbb{R}^N \to \mathbb{R}$ o funcţie derivabilă cu proprietăţile:*

1. *$\exists \alpha > 0$ astfel ca $(\nabla J(v) - \nabla J(w), v - w) \geq \alpha \|v - w\|^2$ pentru orice $v, w \in \mathbb{R}^N$*
2. *$\exists M > 0$ astfel ca $\|\nabla J(v) - \nabla J(w)\| \leq M\|v - w\|$ pentru orice $v, w \in \mathbb{R}^N$.*

*Dacă există două numere $a$ şi $b$ astfel ca*

$$0 < a \leq r_k \leq b < \frac{2\alpha}{M^2}, \quad \forall k \geq 0, \tag{9.18}$$

*atunci şirul dat de metoda gradientului (9.3)-(9.7) cu paşii variabili $r_k$ converge la minimul unic $x^*$ al lui $J$ şi există o constantă $\beta = \beta(\alpha, M, a, b) < 1$ astfel ca*

$$\|x^k - x^*\| \leq \beta^k \|x^0 - x^*\|. \tag{9.19}$$

*Proof.* Folosind faptul că $\nabla J(x^*) = 0$ avem că

$$x^{k+1} - x^* = x^k - x^* - r_k(\nabla J(x^k) - \nabla J(x^*))$$

şi prin urmare

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 + (r_k)^2 \|\nabla J(x^k) - \nabla J(x^*)\|^2 - 2r_k(\nabla J(x^k) - \nabla J(x^*), x^k - x^*) \leq$$

$$\leq (1 + M^2(r_k)^2 - 2\alpha r_k)\|x^k - x^*\|^2.$$

Trinomul $\tau(r) = 1 - 2\alpha r + M^2 r^2$ are pe intervalul $(0, \frac{2\alpha}{M^2})$ valori strict mai mici ca 1 şi

$$\tau(r) = 1 - 2\alpha r + M^2 r^2 \leq \max\{\tau(a), \tau(b)\} := \beta^2 < 1, \quad \forall r \in (a, b).$$

Prin urmare, obţinem că

$$\|x^{k+1} - x^*\| \leq \beta \|x^k - x^*\|,$$

iar prin iterare

$$\|x^{k+1} - x^*\| \leq \beta^k \|x^0 - x^*\|.$$

Demonstraţia este încheiată. □

**Observaţia 9.2.** *Dacă $J$ este de două ori derivabilă condiţia 2. este verificată dacă*

$$\sup_{v \in \mathbb{R}^N} \|\nabla^2 J(v)\| \leq M. \tag{9.20}$$

**Observaţia 9.3.** *În cazul (9.17) condiţiile 1. şi 2. din teorema precedentă sunt verificate:*

$$\alpha = \lambda_1$$

*şi*

$$\nabla^2 J(v) = A \Rightarrow \sup_{v \in \Omega} \|\nabla^2 J(v)\| = \|A\| = \rho(A) = \lambda_N := M.$$

*Prin urmare, metoda gradientului converge pentru orice paşi $r_k$ care verifică*

$$a \leq r_k \leq \frac{2\lambda_1}{\lambda_N^2}, \quad \forall k \geq 0.$$

*Să observăm că pasul optimal nu verifică restricţia:*

$$r_k = \frac{\|\nabla J(x^k)\|^2}{(A\nabla J(x^k), \nabla J(x^k))} \in \left(\frac{1}{\lambda_N}, \frac{1}{\lambda_1}\right) \not\subseteq \left(a, \frac{2\lambda_1}{\lambda_N^2}\right).$$

*Să mai notăm că, de fapt, se poate demonstra convergenţa pentru orice $\lambda_1 \leq r_k \leq \frac{2}{\lambda_N}$.*

**Observaţia 9.4.** *Metoda gradientului este un caz particular al metodei Newton modificată pentru găsirea zerourilor lui $\nabla J$,*

$$x^{k+1} = x^k - (\nabla^2 J(x^k))^{-1} \nabla J(x^k),$$

*în care $(\nabla^2 J(x^k))^{-1}$ se înlocuie prin matrice mai simplă $r_k I$, cu paşii $r_k$ constanţi, variabili sau optimali.*

# Cursul 10

# Probleme de optimizare
# Metoda gradientului conjugat

În acest caz vom considera funcţionala $J : \mathbb{R}^N \to \mathbb{R}$ de forma particulară,

$$J(x) = \frac{1}{2}(Ax, x) - (b, x) \tag{10.1}$$

unde $A \in \mathcal{M}_N(\mathbb{R})$ este o matrice simetrică şi definită pozitiv iar $b \in \mathbb{R}^N$ este un vector. Se ştie că în acest caz $J$ este eliptică şi are un unic punct de minim $x^* \in \mathbb{R}^N$,

$$J(x^*) = \min_{x \in \mathbb{R}^N} J(x). \tag{10.2}$$

În plus, $x^*$ este unicul element cu proprietatea că $\nabla J(x^*) = 0$.

Pentru aproxiarea lui $x^*$ cu **metoda gradientului conjugat**, alegem $x^0 \in \mathbb{R}^N$ şi definim

$$x^{k+1} = x^k + p^k \qquad k \geq 0 \tag{10.3}$$

unde direcţia de coborâre $p^k$ se aleg astfel încât

$$J(x^{k+1}) = J(x^k + p^k) = \min_{x \in G_k} J(x^k + x) \qquad k \geq 0 \tag{10.4}$$

unde $G_k = \text{Span}\{\nabla J(x^0), \nabla J(x^1), ..., \nabla J(x^k)\}$.

Mai întâi să stabilim existenţa şi unicitatea vectorului $p^k$ precum şi o caracterizare a lui.

**Lema 10.1.** *Fie $x^0 \in \mathbb{R}^N$. Pentru fiecare $k \geq 0$, există un unic vector $p^k \in G_k$ cu proprietatea că*

$$x^{k+1} = x^k + p^k \text{ şi } J(x^{k+1}) = \min_{x \in G_k} J(x^k + x), \tag{10.5}$$

*unde $G_k = Span\{\nabla J(x^0), \nabla J(x^1), ..., \nabla J(x^k)\}$. În plus, au loc relaţiile*

$$\Big(\nabla J(x^k + p^k), \nabla J(x^j)\Big) = 0, \qquad 0 \leq j \leq k. \tag{10.6}$$

*Proof.* Considerăm aplicaţia $x \in \mathbb{R}^N \longrightarrow J(x^k + x) \in \mathbb{R}$ care este eliptică. Spaţiul $G_k$ este o mulţime închisă, convexă şi nevidă. Conform Teoremei **??**, există un unic $p^k \in G_k$ care verifică (10.5). În plus, se verifică

$$\left(\nabla J(x^k + p^k), v - p^k\right) \geq 0, \qquad \forall v \in G_k,$$

ceea ce, ţinând cont că $G_k$ este un spaţiu liniar, implică

$$\left(\nabla J(x^k + p^k), v\right) = 0, \qquad \forall v \in G_k. \tag{10.7}$$

Demonstraţia lemei este completă. □

**Observaţia 10.1.** *Conforma lui (10.5), pentru fiecare $k \geq 1$, avem că*

$$x^k \in x^0 + G_{k-1}. \tag{10.8}$$

*Cum $G_k \subseteq G_{k-1}$, rezultă că*

$$J(x^{k+1}) = \min_{x \in G_k} J(x^k + x) = \min_{x \in G_k} J(x^0 + x) \tag{10.9}$$

*şi şirul $(J(x^k))_{k \geq 0}$ este descrescător.*

**Observaţia 10.2.** *Dacă $x^{k+1}$, $k \geq 0$, este dat de (10.5) atunci există $k \in \{0, 1, ..., N-1\}$ astfel încât $x^k$ este minimizatorul lui $J$ în $\mathbb{R}^N$. Într-adevăr, dacă $\nabla J(x^k) = 0$ pentru un $k \leq N-1$, atunci $x^k$ este minimizatorul unic al funcţiei (strict) convexe $J$ în $\mathbb{R}^N$. Dacă $\nabla J(x^k) \neq 0$ pentru orice $0 \leq k \leq N-1$, atunci*

$$J(x^N) := J(x^{N-1} + p^{N-1}) = \min_{x \in G_{N-1}} J(x^{N-1} + x). \tag{10.10}$$

*Din relaţiile (10.6) obţinem că $G_{N-1}$ este un subspaţiu de dimensiune $N$ al lui $\mathbb{R}^N$, de unde deducem că $G_{N-1} = \mathbb{R}^N$. Din (10.10) obţinem că $x^N$ este minimizatorul unic al lui $J$ în $\mathbb{R}^N$. Prin urmare, metoda gradientului conjugat converge în cel mult $N$ iteraţii.*

**Observaţia 10.3.** *Dacă $\nabla J(x^j) \neq 0$ pentru $0 \leq j \leq k$, atunci $\left(p^k, \nabla J(x^k)\right) \neq 0$. Într-adevăr, dacă $\left(p^k, \nabla J(x^k)\right) = 0$, atunci $p^k \in G_{k-1}$. Cum*

$$J(x^{k+1}) = \min_{x \in G_k} J(x^0 + x) = J(x^k + p^k) = \min_{x \in G_{k-1}} J(x^0 + x) = J(x^k)$$

*rezultă că $x^{k+1} = x^k$ şi din (10.6) deducem $\nabla J(x^k) = 0$ ceea ce reprezintă o contradicţie.*

Următorea proprietate caracterizează metoda gradientului conjugat şi îi dă şi numele.

**Lema 10.2.** *Presupunem că $\nabla J(x^j) \neq 0$ pentru $0 \leq j \leq k$. Atunci, pentru orice $0 \leq l, m \leq k$, $l \neq m$, avem că*

$$(Ap^l, p^m) = 0. \tag{10.11}$$

*Proof.* Fără a restrânge generalitatea, putem presupune că $l < m$. Avem că

$$(Ap^l, p^m) = (p^l, Ap^m) = \left(p^l, A(x^{m+1} - x^m)\right) = \left(p^l, \nabla J(x^{m+1}) - \nabla J(x^m)\right).$$

Cum $l < m$ şi $p^l \in G_l$, folosind relaţiile (10.6), deducem că

$$\left(p^l, \nabla J(x^{m+1})\right) = \left(p^l, \nabla J(x^m)\right) = 0$$

şi demonstaţia se încheie. □

**Observaţia 10.4.** *Relaţiile (10.11) reprezintă o ortogonalitate a direcţiilor de coborâre în raport cu produsul scalar $(A\,\cdot\,,\,\cdot\,)$. Se obişnuieşte să se spună că direcţiile de coborâre sunt "conjugate". De aici provine şi numele metodei.*

Vom determina acum direcţiile $p^k$. Avem că

$$p^k = \sum_{j=0}^{k} r_j^k \nabla J(x^j).$$

Folosind (10.11), pentru orice $l < k$,

$$0 = (p^k, Ap^l) = \sum_{j=0}^{k} r_j^k \left(\nabla J(x^j), Ap^l\right) = \sum_{j=0}^{k} r_j^k \left(\nabla J(x^j), \nabla J(x^{l+1}) - \nabla J(x^l)\right)$$

de unde

$$r_{l+1}^k \|\nabla J(x^{l+1})\|^2 - r_l^k \|\nabla J(x^l)\|^2 = 0 \Rightarrow r_l^k = r_{l+1}^k \frac{\|\nabla J(x^{l+1})\|^2}{\|\nabla J(x^l)\|^2}.$$

Prin urmare,

$$r_j^k = r_k^k \frac{\|\nabla J(x^k)\|^2}{\|\nabla J(x^j)\|^2} \qquad 0 \le j \le k \tag{10.12}$$

de unde obţinem că

$$p^k = r_k^k \sum_{j=0}^{k} \frac{\|\nabla J(x^k)\|^2}{\|\nabla J(x^j)\|^2} \nabla J(x^j). \tag{10.13}$$

Să notăm

$$d^k = -\nabla J(x^k) - \sum_{j=0}^{k-1} \frac{\|\nabla J(x^k)\|^2}{\|\nabla J(x^j)\|^2} \nabla J(x^j)$$

şi să observăm că

$$d^k = -\nabla J(x^k) + \frac{\|\nabla J(x^k)\|^2}{\|\nabla J(x^{k-1})\|^2} d^{k-1}. \tag{10.14}$$

Pentru completa determinare a lui $p^k = -r_k^k d^k$ mai ramâne de calculat $r_k^k$.
Din relaţia $\left(\nabla J(x^{k+1}), \nabla J(x^k)\right) = 0$ rezultă că

$$0 = \left(Ax^{k+1} - b, \nabla J(x^k)\right) = \left(Ax^k - b + Ap^k, \nabla J(x^k)\right) = \left(\nabla J(x^k) - r_k^k Ad^k, \nabla J(x^k)\right) =$$

$$= \left(\nabla J(x^k) - r_k^k Ad^k, \frac{\|\nabla J(x^k)\|^2}{\|\nabla J(x^{k-1})\|^2} d^{k-1} - d^k\right) = \left(\nabla J(x^k) - r_k^k Ad^k, -d^k\right)$$

de unde rezultă că

$$r_k^k \left(Ad^k, d^k\right) = \left(\nabla J(x^k), d^k\right)$$

şi prin urmare

$$r_k^k = \frac{\left(\nabla J(x^k), d^k\right)}{(Ad^k, d^k)} = -\frac{\left(\nabla J(x^k), \nabla J(x^k)\right)}{(Ad^k, d^k)} := -r_k. \tag{10.15}$$

Avem următorul algoritm (**gradient conjugat**):

1. Se alege $x^0$ şi $\varepsilon > 0$
2. $g^0 = Ax^0 - b$
3. $d^0 = -g^0$
4. $r_0 = \dfrac{(g^0, g^0)}{(Ad^0, d^0)}$
5. $x^1 = x^0 + r_0 d^0$
6. $g^1 = Ax^1 - b = g^0 + r_0 Ad^0$
7. Pentru $k = 1, 2, ...N$ (sau cât timp $\|g^k\| > \varepsilon$)
   - 7.1. $d^k = -g^k + \frac{(g^k, g^k)}{(g^{k-1}, g^{k-1})} d^{k-1}$
   - 7.2. $r_k = \dfrac{(g^k, g^k)}{(Ad^k, d^k)}$
   - 7.3. $x^{k+1} = x^k + r_k d^k$
   - 7.4. $g^{k+1} = Ax^{k+1} - b = g^k + r_k Ad^k$.

**Observaţia 10.5.** *Cele mai multe operaţii în metoda gradientului conjugat se realizează la calculul produselor $Ax^k$ şi $Ad^k$, unde sunt necesare $N^2$ înmulţiri şi $N^2$ adunări. Prin urmare s-ar părea că numărul total de operaţii va fi de ordin $2N^3$. De fapt, în practică, se realizează mult mai puţine operaţii deoarece*

- *Matricea A este rară (cu foarte multe elemente zero) prin urmare produsele şi sumele sunt mult mai puţin numeroase.*
- *Nu se realizează niciodată N iteraţii. Algoritmul se poate opri, cu un rezultat foarte bun, mult mai repede.*

## Exerciţii

1. Pentru funcţionala

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

construiţi primele două aproximaţii cu metoda gradientului conjugat luând $\begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Care este punctul de minim al lui $f$?

2. Pentru funcţionala $f(x_1, x_2) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2$, construiţi primele două aproximaţii cu metoda gradientului conjugat luând $x_1^0 = x_2^0 = 0$.

3. Fie $n \in \mathbb{N}^*$. Arătaţi că vectorii $s^k$, $k = 1, 2, ..., n$, definiţi prin

$$s_i^k = \begin{cases} i & \text{dacă } i \leq k \\ 0 & \text{în caz contrar} \end{cases}$$

sunt conjugaţi în raport cu matricea tridiagonală $G \in \mathcal{M}_n(\mathbb{R})$, definită de $G_{ii} = 2$, $G_{i,i+1} = G_{i+1,i} = -1$ pentru orice $i$.

4. Consideraţi minimizarea funcţionalei $f : \mathbb{R}^4 \to \mathbb{R}$, dată de $f(x) = \frac{1}{2}x^tGx + b^tx$ unde

$$G = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}, b = \begin{pmatrix} -1 \\ 0 \\ 2 \\ \sqrt{5} \end{pmatrix}$$

Aplicaţi metoda gradientului conjugat acestei probleme cu $x^0 = (0, 0, 0, 0)^t$ şi demonstraţi că este convergentă în două iteraţii. Verificaţi că există doar doi vectori liniar independenţi în şirul $\nabla f(x^0), G\nabla f(x^0), G^2\nabla f(x^0), ....$