# TEHNICI DE OPTIMIZARE
# Curs 10

Andrei Pătraşcu

Departament Informatică
Universitatea din Bucureşti

- Examenul s-a stabilit pe 16.05.2025 (ora 9)
- Examenul va avea loc cu materiale la dispoziţie
- Materialele vor fi doar fizice, nimic electronic
- Examenul va dura 2h

- **Algoritmi pentru (POCi) convexe.**
- Recapitulare:
  - Mulţimi şi funcţii convexe.
  - Algoritmi pentru (POfC).
  - Algoritmi pentru (POCi) .
  - Dualitate şi optimalitate.

QP dimensiune $n$:

$$\min_{x} \quad \frac{1}{2}x^T Hx + c^T x \ \text{ s.l. } \ Ax \leq b.$$

Problema este convexă $H \succ 0$, $m$ constrângeri liniare $\Rightarrow$ dualitate tare

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^T Hx + c^T x + \lambda^T(Ax - b)$$

$$\phi(\lambda) = \min_{x} \quad \frac{1}{2}x^T Hx + (c + A^T\lambda)^T x - \lambda^T b$$

Kuhn-Tucker:

$$
\begin{aligned}
& Hx + c + A^T\lambda = 0 \\
& \lambda_i(A_i x - b_i) = 0 \quad \forall i = 1, \cdots, m \\
& Ax \le b, \lambda \ge 0.
\end{aligned}
$$

Forma soluţiei primale:

$$
\begin{aligned}
x(\lambda) &= H^{-1}(c + A^T\lambda) \\
\phi(\lambda) &= -\frac{1}{2}(c + A^T\lambda)^T H^{-1}(c + A^T\lambda) - \lambda^T b.
\end{aligned}
$$

**Problema duală**: QP dimensiune *m*

$$
\max_{\lambda \ge 0} \; -\frac{1}{2}(c + A^T\lambda)^T H^{-1}(c + A^T\lambda) - \lambda^T b
$$

$$\min_x \; f(x) \; \text{ s.l. } \; g(x) \leq 0.$$

Metoda Gradientului Dual:

$$x^{k+1} = \arg\min_x \mathcal{L}(x, \lambda^k)$$
$$\lambda^{k+1} = \pi_{\geq 0}(\lambda^k + \alpha g(x^{k+1}))$$

$$\min_x \; f(x) \;\; \text{s.l.} \;\; g(x) \leq 0.$$

Problema duală:

$$\max_{\lambda \geq 0} \; \phi(\lambda), \qquad \phi(\lambda) = \min_x \mathcal{L}(x, \lambda)$$

- notăm $x(\lambda) = \arg\min_x \mathcal{L}(x, \lambda)$
- Dacă $f$ este $\sigma-$tare convexă, atunci $\nabla \phi(\lambda) = g(x(\lambda))$ este $\frac{1}{\sigma}-$continuu Lipschitz

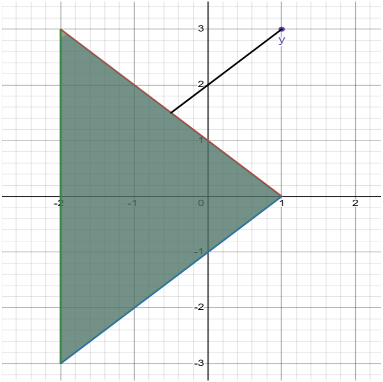
$$\min_x \; f(x) \;\; \text{s.l.} \;\; g(x) \leq 0.$$

Metoda Gradientului Dual:

$$\lambda^{k+1} = \pi_{\geq 0}(\lambda^k + \alpha \nabla \phi(\lambda^k))$$

- Iteraţia este identică cu MGP pentru problema duală!
- $\lambda^k \rightarrow \lambda^* \Rightarrow x(\lambda^k) \rightarrow x^*$

$$\min_{x} \quad \|x - y\|_2^2 \;\text{ s.l. }\; Ax \leq b.$$



$$\min_{x} \quad \|x - y\|_2^2 \text{ s.l. } Ax \leq b.$$

$$y = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 0 \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$\mathcal{L}(x, \lambda) = \quad \|x - y\|_2^2 + \lambda^T (Ax - b)$$

$$\phi(\lambda) = \min_{x} \mathcal{L}(x, \lambda)$$

Problema duală: $\max\limits_{\lambda \geq 0} \; -\frac{1}{4}\|A^T \lambda\|_2^2 + \lambda^T (Ay - b)$

Problema proiecţiei ortogonale:

$$\min_x \ \|x - y\|_2^2 \ \text{ s.l. } \ Ax \leq b.$$

$$\nabla\phi(\lambda) = -\frac{1}{2}AA^T\lambda + (Ay - b)$$

Metoda Gradientului Dual:

$$\lambda^{k+1} = \pi_{\geq 0}\left([I - \alpha\frac{1}{2}AA^T]\lambda^k + \alpha(Ay - b)\right)$$

## Exerciţiu

Pornind din $\lambda^0 = [1\ 1\ 1]$, calculaţi un pas de MGD cu pas constant $\alpha = \frac{1}{L}$.

- Algoritmi pentru (POCi) convexe.
- **Recapitulare:**
  - Mulţimi şi funcţii convexe.
  - Algoritmi pentru (POfC).
  - Algoritmi pentru (POCi) .
  - Dualitate şi optimalitate.

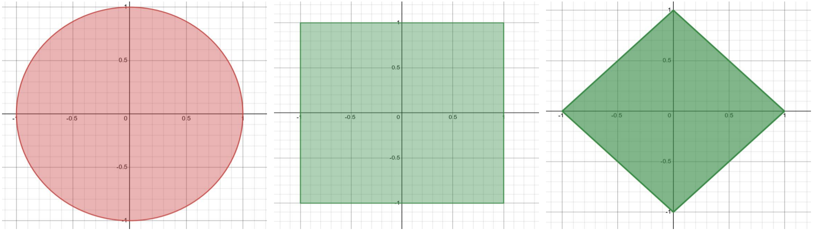
## Definiţie

*Mulţimea Q este **convexă** dacă şi numai dacă* $\forall x, y \in Q$

$$\alpha x + (1-\alpha)y \in Q \qquad \forall \alpha \in [0,1].$$

*Mai mult, Q este **închisă** dacă îşi conţine toate punctele limită ale şirurilor din Q.*

Pe scurt: segmentul aflat între oricare două puncte ale mulţimii se află, de asemenea, în mulţime.



www.desmos.com

Fie $C_1, C_2, \cdots, C_m \subset \mathbb{R}^n$ mulţimi convexe închise.

- *Intersecţia*: mulţimea $C = \bigcap_{i=1}^{m} C_i$ este închisă şi convexă

Fie $C_1, C_2, \cdots, C_m \subset \mathbb{R}^n$ mulţimi convexe închise.

- *Intersecţia*: mulţimea $C = \bigcap_{i=1}^{m} C_i$ este închisă şi convexă
- *Produs cartezian*: mulţimea $C = C_1 \times C_2 \times \cdots \times C_m$ este închisă şi convexă.

Fie $C_1, C_2, \cdots, C_m \subset \mathbb{R}^n$ mulţimi convexe închise.

- *Intersecţia*: mulţimea $C = \bigcap_{i=1}^{m} C_i$ este închisă şi convexă
- *Produs cartezian*: mulţimea $C = C_1 \times C_2 \times \cdots \times C_m$ este închisă şi convexă.
- *Compunerea cu un operator afin*: fie $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, mulţimea $C_i' = \{z \in C_i : Az - b\}$ este închisă şi convexă.

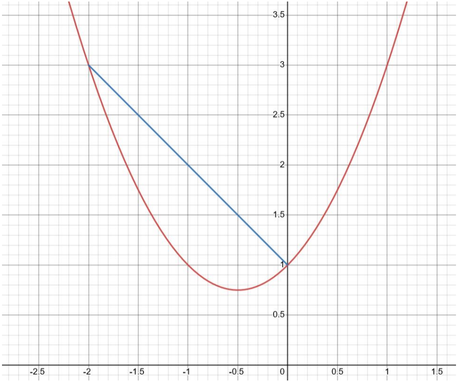
- $\{x \in \mathbb{R}^2 : -1 \leq x_1 + x_2 - 2 \leq 1\}$

- $\{x \in \mathbb{R}^2 : -1 \leq x_1 + x_2 - 2 \leq 1\}$
- $\{x \in \mathbb{R}^2 : -1 \leq x_1^2 + x_2 \leq 1\}$

- $\{x \in \mathbb{R}^2 : -1 \leq x_1 + x_2 - 2 \leq 1\}$
- $\{x \in \mathbb{R}^2 : -1 \leq x_1^2 + x_2 \leq 1\}$
- $\{x \in \mathbb{R}^2 : x \geq 0, x_1^2 - x_2^2 \leq 1\}$

Funcţia scalară $f : \mathbb{R}^n \to (-\infty, +\infty]$ este convexă dacă:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \quad \forall \alpha \in [0,1]$$




www.desmos.com

Fie $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ funcţii convexe închise.

- *Suma*: pentru $\alpha \in \mathbb{R}^m_+$, funcţia $f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$ este închisă şi convexă

Fie $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ funcţii convexe închise.

- *Suma*: pentru $\alpha \in \mathbb{R}^m_+$, funcţia $f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$ este închisă şi convexă
- *Maximum*: funcţia $f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}$ este închisă şi convexă.

# Operaţii care păstrează convexitatea

Fie $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ funcţii convexe închise.

- *Suma*: pentru $\alpha \in \mathbb{R}^m_+$, funcţia $f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$ este închisă şi convexă
- *Maximum*: funcţia $f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}$ este închisă şi convexă.
- *Compunerea interioară cu un operator afin*: fie $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, funcţia $g(x) = f(Ax - b)$ este închisă şi convexă.

Fie $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ funcţii convexe închise.

- *Suma*: pentru $\alpha \in \mathbb{R}^m_+$, funcţia $f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$ este închisă şi convexă
- *Maximum*: funcţia $f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}$ este închisă şi convexă.
- *Compunerea interioară cu un operator afin*: fie $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, funcţia $g(x) = f(Ax - b)$ este închisă şi convexă.
- *Compunerea exterioară cu o funcţie crescătoare*: fie $\phi : \mathbb{R} \to \mathbb{R}$ o funcţie crescătoare, funcţia $g(x) = \phi(f(x))$ este închisă şi convexă.

- $\max\{0, (x_1 + x_2)^2 - 1\}$

- $\max\{0, (x_1 + x_2)^2 - 1\}$
- $\min\{1, |x|\}$

# Exerciţii

- $\max\{0, (x_1 + x_2)^2 - 1\}$
- $\min\{1, |x|\}$
- $|x_1 + x_2|^5 + 2^{x_1 - x_2}$

- $\max\{0, (x_1 + x_2)^2 - 1\}$
- $\min\{1, |x|\}$
- $|x_1 + x_2|^5 + 2^{x_1 - x_2}$
- $\log(1/x)$

- $\max\{0, (x_1 + x_2)^2 - 1\}$
- $\min\{1, |x|\}$
- $|x_1 + x_2|^5 + 2^{x_1 - x_2}$
- $\log(1/x)$
- $\log(1/(c^T x))$

- Algoritmi pentru (POCi) convexe.
- Mulţimi şi func ctii convexe.
- **Algoritmi pentru (POfC)**.
- Algoritmi pentru (POCi) .
- Dualitate şi optimalitate.

$$(POfC): \qquad \min_x \; f(x)$$

### Teoremă (Fermat)

*Presupunem funcţia $f$ diferenţiabilă. Fie $x^*$ un punct de minim al $f(\cdot)$ din $\mathbb{R}^n$, atunci: $\nabla f(x^*) = 0$.*

### Teoremă

*Presupunem funcţia $f$ este dublu diferenţiabilă. Fie $x^*$ un punct de minim al $f(\cdot)$ din $\mathbb{R}^n$, atunci: $\nabla^2 f(x^*) \succeq 0$.*

### Teoremă

*Presupunem funcţia $f$ este dublu diferenţiabilă. Dacă $\nabla f(x^*) = 0$ şi:*

$$\nabla^2 f(x^*) \succ 0$$

*atunci $x^*$ este minim local.*

# Exerciţii

- $\min_x \; x_1^2 - x_2^2 + 2x_1x_2$

- $\min_x \; x_1^2 - x_2^2 + 2x_1x_2$
- $\min_x \; x_1^3 - x_2 + x_1x_2$

## Definiţie

*O funcţie* $f \in \mathcal{S}^{\sigma}$ *este* $\sigma-$*tare convexă dacă satisface:*

- *Ordin* 0*:*
  $f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)-\alpha(1-\alpha)\frac{\sigma}{2}\|x-y\|^2, \ \ \forall\alpha \in [0,1], x, y$
- *Ordin* 1*:* $f(x) \geq f(y) + \nabla f(y)^T(x-y) + \frac{\sigma}{2}\|x-y\|^2, \ \ \forall x, y$
- *Ordin* 2*:* $\nabla^2 f(x) \succeq \sigma I_n \ \ \forall x$

Exemple:

- $x \to (c^T x)^2 + \frac{1}{2}\|x\|_2^2$
- $x \to \|Ax - b\|_2^2 + \|x\|_2^2$
- $x \to \frac{1}{2}x^T Hx + q^T x, \ \ H \succ 0$

## Metoda Gradient

**Algorithm 1:** Metoda Gradient $(x^0, \epsilon, \{\alpha_k\}_{k\geq 0})$:

**Data:** $k := 0$

**while** criteriu oprire = fals **do**

  Calculează: $\nabla f(x^k)$

  $x^{k+1} = x^k - \alpha_k \nabla f(x^k)$

  $k := k + 1$

**end**

- pas constant: $\alpha_k = \alpha$
- cea mai abruptă pantă: $\alpha_k = \arg\min_\alpha f(x^k - \alpha \nabla f(x^k))$
- adaptiv

## Teoremă

*Fie f diferenţiabilă pe* $\mathbb{R}^n$ *cu gradientul* $\nabla f$ *continuu Lipschitz:*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

*De asemenea, presupunem* $\min_x f(x) > -\infty$ *şi* $0 < \alpha < \frac{1}{2L}$. *Atunci şirul generat de Metoda Gradient* $x^{k+1} = x^k - \alpha\nabla f(x^k)$ *satisface:*

$$\lim_{k\to\infty} \nabla f(x^k) = 0,$$

*şi* $f(x^{k+1}) \leq f(x^k)$.

## Teoremă (Rată de convergenţă (convexitate))

*Fie f convexă cu gradientul* $\nabla f$ *continuu Lipschitz:*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

*De asemenea, presupunem* $\min_x f(x) > -\infty$ *şi* $0 < \alpha < \frac{1}{2L}$. *Atunci şirul generat de Metoda Gradient* $x^{k+1} = x^k - \alpha\nabla f(x^k)$ *satisface:*

$$f(x^k) - f^* \leq \frac{L\|x^0 - x^*\|^2}{2k} \quad \forall k \geq 0.$$

Aproximarea pătratică în $x^k$

$$f(x) \approx f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T H_k (x - x^k)$$

Alegerea $H_k = \alpha I$ stă la baza MG.

Dacă funcţia este dublu diferenţiabilă atunci calitatea maximă aproximării se obţine prin alegerea:

$$H_k := \nabla^2 f(x^k).$$

O nouă iteraţie (Newton):

$$x^{k+1} := \arg\min_{x} \; f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k)$$

La optim avem:

$$\begin{aligned} \nabla^2 f(x^k)(x^{k+1} - x^k) + \nabla f(x^k) &= 0 \\ \nabla^2 f(x^k)(x^{k+1} - x^k) &= -\nabla f(x^k) \\ x^{k+1} &= x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k) \end{aligned}$$

Fie funcţia pătratică:

$$f(x) = \frac{1}{2}x^T Ax - b^T x, \quad A \succ 0.$$

Metoda Newton converge într-un singur pas!

$$\begin{aligned} x^1 &= x^0 - [\nabla^2 f(x^0)]^{-1} \nabla f(x^0) \\ &= x^0 - A^{-1}(Ax^0 - b) = A^{-1}b =: x^* \end{aligned}$$

Cu cât $f$ este mai aproape de o funcţie pătratică, cu atât MN converge mai rapid.

**Algorithm 2:** Metoda Newton $(x^0, \epsilon, \{\alpha_k\}_{k\geq 0})$:

**Data:** $k := 0$

**while** criteriu oprire = fals **do**

  Calculează: $d^k = [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$

  $x^{k+1} = x^k - \alpha_k d^k$

  $k := k + 1$

**end**

## Teoremă (Polyak, caz convex)

*Fie $f$ dublu diferenţiabilă şi $\nabla^2 f$ continuu Lipschitz cu constanta $L$:*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|.$$

*De asemenea, presupunem $f$ tare convexă cu constanta $\sigma$. Atunci, dacă iteraţia iniţială satisface:*

$$q = \frac{L}{2\sigma^2}\|\nabla f(x^0)\| < 1,$$

*atunci $x^k$ generat de MN cu pas constant $\alpha_k = 1$ converge pătratic la optimul global $x^*$, i.e.*

$$\|x^k - x^*\| \leq \frac{2\sigma}{L} q^{2^k}$$

1

## Teoremă (Polyak, caz general neconvex)

*Fie $f$ dublu diferenţiabilă într-o vecinuătate $U$ a unui minim local nesingular $x^*$. De asemenea, presupunem $\nabla^2 f$ continuu Lipschitz cu constanta $L$ în $U$. Atunci, există $\delta$ astfel încât pentru:*

$$\|x^0 - x^*\| < \delta,$$

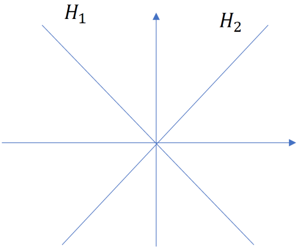
*$x^k$ generat de MN cu $\alpha_k = 1$ converge pătratic la optimul local $x^*$.*

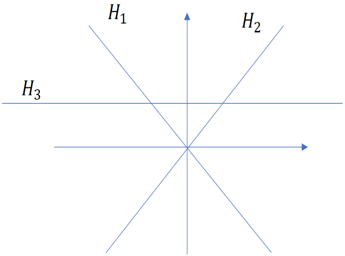
1

Când există o soluţie $\mathbb{E}[a_\xi x - b_\xi] = 0$ avem o problemă de interpolare liniară.

## Interpolare

Definim $C = H_1 \cap \cdots \cap H_m \neq \emptyset$, unde $H_i = \{x : a_i^T x = b_i\}$. Determinaţi un punct din mulţimea $C$, i.e. $x \in C$.



$$(POC): \quad \min_{x \in \mathbb{R}^n} \; f(x) \;\; \text{supus la} \;\; x \in Q$$

- $Q$ mulţime fezabilă (convexă, închisă)

$$(POC): \quad \min_{x \in \mathbb{R}^n} \; f(x) \;\; \text{supus la} \;\; x \in Q$$

- $Q$ mulţime fezabilă (convexă, închisă)
- se urmăreşte determinarea lui $x^* \in Q$ care asigură:

$$f(x^*) \leq f(x), \qquad \forall x \in Q, \|x - x^*\| \leq \epsilon \;\; \text{(minim local)}$$
$$f(x^*) \leq f(x), \qquad \forall x \in Q \;\; \text{(minim global)}$$

$$(POC): \quad \min_{x \in \mathbb{R}^n} \; f(x) \;\; \text{supus la} \;\; x \in Q$$

- $Q$ mulţime fezabilă (convexă, închisă)
- se urmăreşte determinarea lui $x^* \in Q$ care asigură:

$$f(x^*) \leq f(x), \qquad \forall x \in Q, \|x - x^*\| \leq \epsilon \;\; \text{(minim local)}$$
$$f(x^*) \leq f(x), \qquad \forall x \in Q \;\; \text{(minim global)}$$

- echivalent: $\min_x \; f(x) + \imath_Q(x)$, unde $\imath_Q(x) = \begin{cases} 0 & x \in Q \\ \infty & x \notin Q \end{cases}$

Numim mulţimea fezabilă simplă dacă următoarele obiecte pot fi calculate eficient:

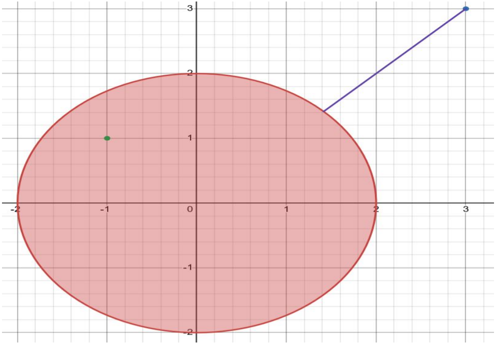
- punct fezabil: $x \in Q$
- proiecţia ortogonală: $\pi_Q(y) := \mathrm{argmin}_x \ \ \|x - y\|$ s.l. $x \in Q$
- oracol liniar: $\phi_Q(y) := \mathrm{argmin}_x \ \ x^T y$ s.l. $x \in Q$
- mulţimi active
- ...

# O problemă particulară

Problema proiecţiei ortogonale Euclidiene:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|x - y\|_2^2$$
$$\text{s.l. } x \in Q$$

Soluţia $\pi_Q(y)$: punctul cel mai apropiat de $y$, care aparţine mulţimii fezabile $Q$



Problema proiecţiei ortogonale Euclidiene:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|x - y\|_2^2$$
$$\text{s.l. } x \in Q$$

Soluţia $\pi_Q(y)$: punctul cel mai apropiat de $y$, care aparţine mulţimii fezabile $Q$

- Dacă $Q$ este închisă atunci $\pi_Q(\cdot)$ există
- $y \in Q \rightarrow \pi_Q(y) = y$
- Condiţii de optimalitate: $(y - \pi_Q(y))^T(x - \pi_Q(y)) \leq 0 \quad \forall x \in Q$.

# O problemă particulară

- bila $Q := B(0, r) = \{x \in \mathbb{R}^n : \ \|x\| \leq r\} \Rightarrow \pi_Q(y) = \left\{1, \frac{r}{\|y\|}\right\} y$

- bila $Q := B(0,r) = \{x \in \mathbb{R}^n : \ \|x\| \leq r\} \Rightarrow \pi_Q(y) = \left\{1, \frac{r}{\|y\|}\right\} y$
- hiperparalelipiped
  $Q := [l,u]^n = \{x \in \mathbb{R}^n : \ x \in [l,u]\} \Rightarrow \pi_Q(y)_i = \max\{l_i, \min\{u_i, y_i\}\}$

## O problemă particulară

- bila $Q := B(0,r) = \{x \in \mathbb{R}^n : \ \|x\| \leq r\} \Rightarrow \pi_Q(y) = \left\{1, \frac{r}{\|y\|}\right\} y$
- hiperparalelipiped
  $Q := [l,u]^n = \{x \in \mathbb{R}^n : \ x \in [l,u]\} \Rightarrow \pi_Q(y)_i = \max\{l_i, \min\{u_i, y_i\}\}$
- hiperplan $Q := H(a,b) = \{x \in \mathbb{R}^n : \ a^T x = b\} \Rightarrow \pi_Q(y) = y - \frac{a^T y - b}{\|a\|_2^2} a$

# O problemă particulară

- bila $Q := B(0, r) = \{x \in \mathbb{R}^n : \ \|x\| \leq r\} \Rightarrow \pi_Q(y) = \left\{1, \frac{r}{\|y\|}\right\} y$
- hiperparalelipiped
  $Q := [l, u]^n = \{x \in \mathbb{R}^n : \ x \in [l, u]\} \Rightarrow \pi_Q(y)_i = \max\{l_i, \min\{u_i, y_i\}\}$
- hiperplan $Q := H(a, b) = \{x \in \mathbb{R}^n : \ a^T x = b\} \Rightarrow \pi_Q(y) = y - \frac{a^T y - b}{\|a\|_2^2} a$
- semispaţiu
  $Q := S(a, b) = \{x \in \mathbb{R}^n : \ a^T x \leq b\} \Rightarrow \pi_Q(y) = y - \frac{\max\{0, a^T y - b\}}{\|a\|_2^2} a$

$$(POC): \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s. l. } \; x \in Q.$$

Dacă $Q = \mathbb{R}^n$ atunci în toate punctele staţionare $\nabla f(x^*) = 0$.

## Definiţie

*Fie $f$ diferenţiabilă, $x \in Q$ şi $\alpha > 0$ atunci denumim* ***operator de gradient redus peste mulţimea Q*** *transformarea:*

$$\mathcal{G}_Q(x; \alpha) = \frac{1}{\alpha}\left[x - \pi_Q(x - \alpha \nabla f(x))\right]$$

Dacă $Q = \mathbb{R}^n$ atunci $\mathcal{G}_Q(x; \alpha) = \nabla f(x)$.

## Definiţie

*Fie f diferenţiabilă, $x \in Q$ şi $\alpha > 0$ atunci denumim* ***operator de gradient redus peste mulţimea Q*** *transformarea:*

$$\mathcal{G}_Q(x;\alpha) = \frac{1}{\alpha}\left[x - \pi_Q(x - \alpha \nabla f(x))\right]$$

- $f(x) = \frac{1}{2}\|x - y\|^2$

$$\mathcal{G}(x;1) = x - \pi_Q(x - (x - y)) = x - \pi_Q(y) = x - x^*$$

## Definiţie

*Fie f diferenţiabilă, $x \in Q$ şi $\alpha > 0$ atunci denumim* ***operator de gradient redus peste mulţimea Q*** *transformarea:*

$$\mathcal{G}_Q(x;\alpha) = \frac{1}{\alpha}\left[x - \pi_Q(x - \alpha\nabla f(x))\right]$$

- $f(x) = \frac{1}{2}\|x - y\|^2$

$$\mathcal{G}(x;1) = x - \pi_Q(x - (x - y)) = x - \pi_Q(y) = x - x^*$$

- $f(x) = \frac{1}{2}\|Ax - b\|^2, Q = B(0;1)$

$$\begin{aligned}\mathcal{G}(x;1) &= x - \pi_Q(x - A^T(Ax - b)) = x - \pi_Q((I - A^TA)x + A^Tb)\\ &= x - \frac{1}{\|(I - A^TA)x + A^Tb\|}[(I - A^TA)x + A^Tb]\\ &= (1 - \bar{\alpha})x + \bar{\alpha}\nabla f(x).\end{aligned}$$

$$(OPC): \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s. l. } \; x \in Q.$$

Pp. $x \in Q, \mathcal{G}(x) \neq 0$ atunci din definiţia lui $\mathcal{G}$ avem:

$$\begin{aligned}
\alpha^2 \|\mathcal{G}(x;\alpha) - \nabla f(x))\|_2^2 &= \|\pi_Q(x - \alpha \nabla f(x)) - (x - \alpha \nabla f(x))\|_2^2 \\
&= \min_{z \in Q} \|z - (x - \alpha \nabla f(x))\|_2^2 \\
&\leq \|x - (x - \alpha \nabla f(x))\|_2^2 = \alpha^2 \|\nabla f(x)\|_2^2.
\end{aligned}$$

Desfăşurând norma rezultă:

$$\frac{1}{2}\|\mathcal{G}(x;\alpha)\|_2^2 \leq \nabla f(x)^T \mathcal{G}(x;\alpha).$$

$$(OPC): \quad \min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s. l.} \; x \in Q.$$

Dacă $f$ diferenţiabilă atunci:

$$\begin{aligned} f(x - \alpha \mathcal{G}(x;\alpha)) &= f(x) - \alpha \nabla f(x)^T \mathcal{G}(x;\alpha) + o(-\alpha \mathcal{G}(x;\alpha)) \\ &\leq f(x) - \alpha \frac{1}{2} \|\mathcal{G}(x;\alpha)\|_2^2 + o(-\alpha \mathcal{G}(x;\alpha)) \\ &\leq f(x) - \alpha \|\mathcal{G}(x;\alpha)\|_2 \left[ \frac{1}{2} \|\mathcal{G}(x;\alpha)\|_2 - \frac{o(-\alpha \mathcal{G}(x;\alpha))}{\alpha \|\mathcal{G}(x;\alpha)\|_2} \right]. \end{aligned}$$

Un pas $\alpha$ suficient de mic asigură: $f(x - \alpha \mathcal{G}(x;\alpha)) < f(x)$, de aceea orice minim $x^*$ satisface $\mathcal{G}(x^*;\alpha) = 0$.

$$(OPC): \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s. l.} \;\; x \in Q.$$

## Teoremă

*Fie f diferenţiabilă, Q mulţime convexă închisă şi $\alpha$ o constantă pozitivă. Orice minim $x^*$ al problemei OPC satisface:*

$$x^* = \pi_Q(x^* - \alpha \nabla f(x^*)) \;\; \Leftrightarrow \mathcal{G}(x^*; \alpha) = 0$$

*Mai mult, dacă f convexă atunci $x^*$ soluţia OPC dacă şi numai dacă $\mathcal{G}(x^*; \alpha) = 0$.*

- Analog condiţiilor de ordin I pentru cazul neconstrâns (i.e. $Q = \mathbb{R}^n$)

$$(OPC): \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s. l. } \; x \in Q.$$

**Teoremă**

*Fie $f$ diferenţiabilă, $Q$ mulţime convexă închisă şi $\alpha$ o constantă pozitivă. Orice minim $x^*$ al problemei OPC satisface:*

$$x^* = \pi_Q(x^* - \alpha \nabla f(x^*)) \quad \Leftrightarrow \mathcal{G}(x^*; \alpha) = 0$$

*Mai mult, dacă $f$ convexă atunci $x^*$ soluţia OPC dacă şi numai dacă $\mathcal{G}(x^*; \alpha) = 0$.*

- Analog condiţiilor de ordin I pentru cazul neconstrâns (i.e. $Q = \mathbb{R}^n$)
- Condiţii necesare în general, dar nec. şi suficiente în cazul $f$ convexă

$$(OPC): \quad \min_{x \in \mathbb{R}^n} \ f(x)$$
$$\text{s. l. } \ x \in Q.$$

## Teoremă

*Fie f diferenţiabilă, Q mulţime convexă închisă şi $\alpha$ o constantă pozitivă. Orice minim $x^*$ al problemei OPC satisface:*

$$x^* = \pi_Q(x^* - \alpha \nabla f(x^*)) \quad \Leftrightarrow \mathcal{G}(x^*; \alpha) = 0$$

*Mai mult, dacă f convexă atunci $x^*$ soluţia OPC dacă şi numai dacă $\mathcal{G}(x^*; \alpha) = 0$.*

- Analog condiţiilor de ordin I pentru cazul neconstrâns (i.e. $Q = \mathbb{R}^n$)
- Condiţii necesare în general, dar nec. şi suficiente în cazul $f$ convexă
- Cand $\pi_Q$ este tractabil, condiţia de mai sus este verificabilă!

Din relaţiile de mai sus, pentru $\alpha$ suficient de mic obţinem descreşterea:

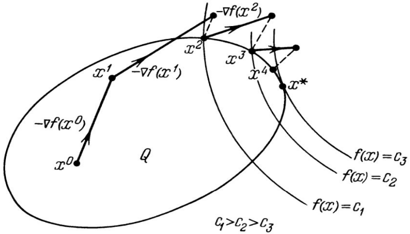
$$f(x - \alpha\mathcal{G}(x)) < f(x)$$

care sugerează iteraţia $x^+ = x - \alpha\mathcal{G}(x)$.

Calcularea operatorlui $\mathcal{G}(\cdot)$ foloseşte $\pi_Q(\cdot)$, care este tractabil doar în anumite cazuri particulare ”simple”: bilă, hipercub, hiperplan etc.

**Metoda Gradientului Proiectat**: iniţial $x^0 \in Q$

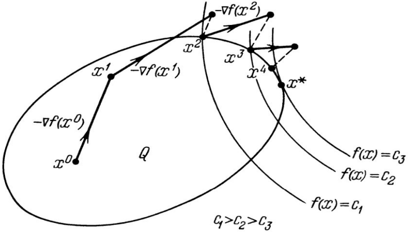
$$x^{k+1} := x^k - \alpha \mathcal{G}(x^k) = \pi_Q(x^k - \alpha \nabla f(x^k))$$




[Polyak, pg. 234]

Interpretare:

$$x^{k+1} = \operatorname*{argmin}_z \; f(x^k) + \nabla f(x^k)^T(z - x^k) + \frac{1}{2\alpha_k}\|z - x^k\|^2$$
$$= \operatorname*{argmin}_z \; \frac{1}{2}\|z - (x^k - \alpha_k \nabla f(x^k))\|^2 \;\; \text{(exercitiu!)}$$




[Polyak, pg. 234]

**Algorithm 3:** Metoda Gradientului Proiectat (MGP) $(x^0, \epsilon, \{\alpha_k\}_{k \geq 0})$:

**Data:** $k := 0$

**while** criteriu oprire = fals **do**

  Calculează: $\nabla f(x^k)$

  $x^{k+1} = \pi_Q(x^k - \alpha_k \nabla f(x^k))$

  $k := k + 1$

**end**

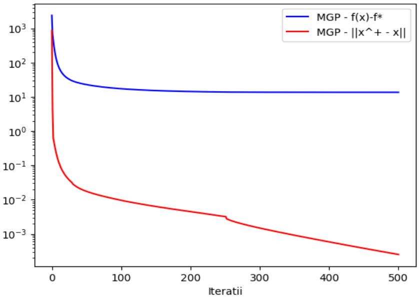
### Teoremă

*Fie $f$ o funcţie diferenţiabilă cu gradient Lipschitz şi $Q$ mulţimea fezabilă convexă. Pentru pasul $\alpha_k = \alpha \in (0, 2/L)$, şirul MGP satisface:*

$$\|\mathcal{G}(x^k)\| \to 0 \;\; \text{când} \;\; k \to \infty.$$

*Mai mult, dacă $f$ convexă şi $X^*$ mulţimea optimă POC, atunci $x^k \to x^* \in X^*$.*

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{2}\|Ax - b\|^2 \ \text{ s.t. } \ l \leq x \leq u.$$



$$\min_{x \in \mathbb{R}^3} \; f(x) = x_1^2 + x_2^2 + x_3^2 + (x_1 + 2x_2 - 4)^2$$
$$\text{s.l.} \; \|x\| \leq 1.$$

Calculaţi prima iteraţie a MGP cu pas constant $\alpha = 1/L$, pornind din $x^0 = [0\ 0\ 0]$.

**Teoremă (Nesterov, 2014)**

*Fie f o funcţie diferenţiabilă cu gradient Lipschitz cu constanta L, tare-convexă cu constanta $\sigma$, iar Q convexă şi închisă. Pentru pasul $\alpha \in \left(0, \frac{2}{L+\sigma}\right]$ avem:*

$$\|x^k - x^*\| \leq \left(1 - \frac{\sigma}{L}\right)^k \|x^0 - x^*\|$$

- Performanţă similară cu cea a MG
- $x^k \in S_f(f^* + \epsilon)$ după $\mathcal{O}\left(\kappa \log(1/\epsilon)\right)$ iteraţii

Ipoteza: Este posibilă calcularea eficientă a operatorului cost liniar:

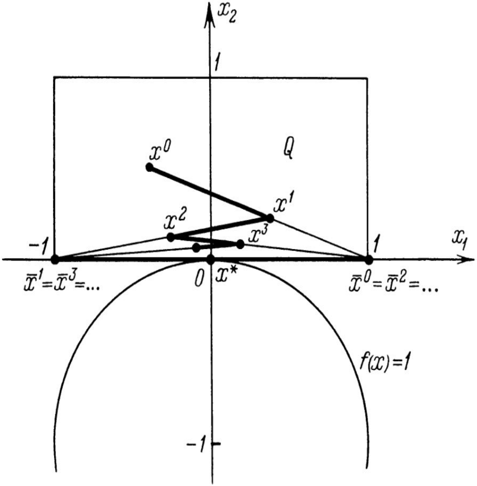
$$\phi_Q(y) := \operatorname*{argmin}_{z \in Q} \; y^T z$$

Ipoteza: Este posibilă calcularea eficientă a operatorului cost liniar:

$$\phi_Q(y) := \operatorname*{argmin}_{z \in Q} \; y^T z$$

Metoda Gradientului Condiţional realizează un model aproximativ liniar al costului şi foloseşte soluţia acestuia la fiecare iteraţie $k$:

$$\begin{aligned} \phi_Q(x^k) :&= \operatorname*{argmin}_{z \in Q} \; \nabla f(x^k)^T z \\ x^{k+1} &= x^k + \alpha_k(\phi_Q(x^k) - x^k). \end{aligned}$$



$x_2$
1
$x^0$
Q
$x^1$
$x^2$
$x^3$
−1
1
$x_1$
$\bar{x}^1=\bar{x}^3=\ldots$
0
$x^*$
$\bar{x}^0=\bar{x}^2=\ldots$
$f(x)=1$
−1

**Algorithm 4:** Metoda Gradientului Condiţional (MGC) $(x^0, \epsilon, \{\alpha_k\}_{k\geq 0})$:

**Data:** $k := 0$

**while** criteriu oprire = fals **do**

  Calculează: $\phi_Q(x^k) := \underset{z\in Q}{\operatorname{argmin}} \; \nabla f(x^k)^T z$

  $x^{k+1} = x^k + \alpha_k(\phi_Q(x^k) - x^k)$

  $k := k + 1$

**end**

## Teoremă

*Fie $f$ o funcţie diferenţiabilă cu gradient Lipschitz şi $Q$ mulţimea fezabilă convexă. Pentru pasul $\alpha_k = \operatorname{argmin}_{\alpha\in[0,1]} f(x^k + \alpha(\phi_Q(x^k) - x^k))$, şirul MGC satisface:*

$$\nabla f(x^k)^T(x^k - \phi_Q(x^k)) \to 0 \ \ \text{când} \ \ k \to \infty.$$

*Mai mult, dacă $f$ convexă atunci $f(x^k) - f^* = \mathcal{O}(1/k)$.*

$$\min_{x \in \mathbb{R}^3} \; f(x) = x_1^2 + x_2^2 + x_3^2 + (x_1 + 2x_2 - 4)^2$$
$$\text{s.l.} \;\; \|x\| \leq 1.$$

Calculaţi prima iteraţie a MGC cu pas constant $\alpha = 1/L$, pornind din $x^0 = [0\ 0\ 0]$.

Calculaţi soluţia următoarelor probleme:

- $\arg\min_{x \in \mathbb{R}^n} \; c^T x \;$ s.l. $\; l \leq x \leq u$
- $\arg\min_{x \in \mathbb{R}^n} \; c^T x \;$ s.l. $\; \|x\| \leq r$

Programare Neliniară: probleme de minimizare supuse la constrângeri de inegalitate

$$(POCi:) \quad \min_{x \in \mathbb{R}^n} \ f(x)$$
$$\text{s.l.} \ \ h(x) \leq 0, g(x) = 0.$$

- Mulţimea fezabilă este definită $Q = \{x \in \mathbb{R}^n : \ h_i(x) \leq 0, g_j(x) = 0, i = 1, \cdots, p; j = 1, \cdots, m\}$.
- Optim-ul global: $f(x^*) \leq f(x), \ \forall x \in Q$.
- POCi convexă dacă $f, h_j$ convexe pentru $j = 1, \cdots, p$ + $g_i$ liniare pentru $i = 1, \cdots, m$.

$$(POCi\ \ convexa:)\quad \min_{x \in \mathbb{R}^n}\ f(x)$$
$$\text{s.l.}\ \ h(x) \leq 0, Ax = b.$$

unde $f, h_i$ funcţii convexe, $h : \mathbb{R}^n \to \mathbb{R}^p, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

$$
\begin{aligned}
(POCi \;\; convexa:) \quad & \min_{x \in \mathbb{R}^n} \; f(x) \\
& \text{s.l.} \;\; h(x) \leq 0, Ax = b.
\end{aligned}
$$

unde $f, h_i$ funcţii convexe, $h : \mathbb{R}^n \to \mathbb{R}^p, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

$$
\begin{aligned}
(SVM:) \quad & \min_{w,b,\xi} \; \frac{1}{2}\|w\|_2^2 + \rho \sum_i \xi_i \\
& \text{s.l.} \;\; y_i(w^T x_i - b) \geq 1 - \xi_i, \xi \geq 0.
\end{aligned}
$$

$$(POCi \;\; convexa:) \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.l. } \; h(x) \leq 0, Ax = b.$$

unde $f, h_i$ funcţii convexe, $h : \mathbb{R}^n \to \mathbb{R}^p, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

$$(Problema \;\; Google:) \quad \min_{x} \; \frac{1}{2}\|Ex - x\|_2^2$$
$$\text{s.l. } \; \sum_i x_i = 1, x \geq 0.$$

$$(POCi:) \quad \min_{x\in\mathbb{R}^n} \; f(x)$$
$$\text{s.l.} \;\; h(x) \leq 0, Ax = b,$$

unde $f, h_i$ funcţii convexe, $h : \mathbb{R}^n \to \mathbb{R}^p, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$.

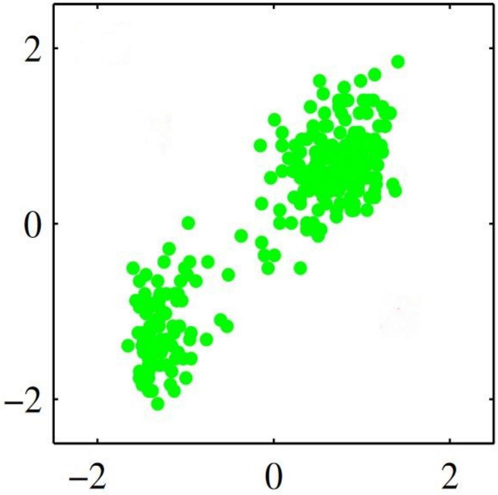
$$(POCi:) \quad \min_{x\in\mathbb{R}^n} \; f(x)$$
$$\text{s.l.} \;\; h(x) \leq 0, Ax = b,$$

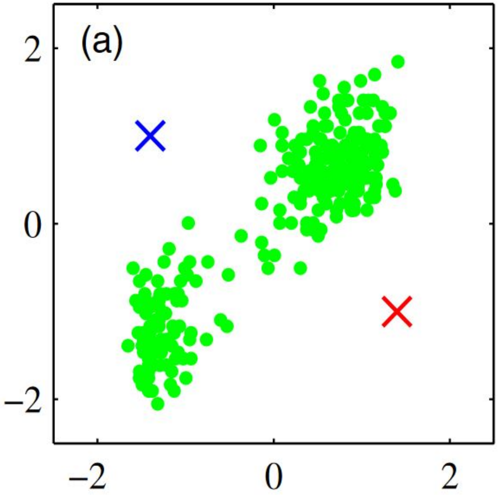
unde $f, h_i$ funcţii convexe, $h : \mathbb{R}^n \to \mathbb{R}^p, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$.

- Funcţia Lagrange $\mathcal{L} : \mathrm{dom}(f) \times \mathbb{R}^m \times \mathbb{R}^p_+$:

$$\mathcal{L}(x, \mu) = f(x) + \mu^T(Ax - b) + \lambda^T h(x)$$

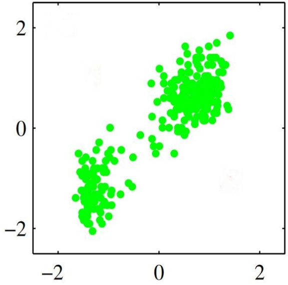
- $\mu, \lambda$ multiplicatori Lagrange (variabile duale)


2
0
−2
−2
0
2


(a)
2
0
−2
−2
0
2

Presupunem setul de date $X = \{x^2, \cdots, x^m\} \subset \mathbb{R}^n$. Definim un grup (cluster) ca un subset de date pentru care distanţele inter-punct sunt mai mici decât distanţele fată de punctele din afara grupului.

**Problema**: Pentru un $K$ dat, partiţionaţi mulţimea $X$ în $K$ grupuri (cluster-e).



- Putem formaliza folosind vectorii $\{\mu^i\}_{i=1}^K$, unde $\mu^i$ reprezintă centrul (prototipul) grupul cu index $i$.
- Soluţie: repartizarea $x^i$ într-un grup şi determinarea prototipului $\mu^i$, astfel încât distanţele euclidiene la $\mu^i$ sa fie minime.

Pentru un punct $x^j$ fixat, repartizarea rezultă din:

$$\min_{1 \leq i \leq K} \|x^j - \mu^i\|^2$$

Problema $K\!-\!Means$:

$$\min_{\mu^1,\cdots,\mu^K} \sum_{j=1}^{m} \min_{1\leq i\leq K} \|x^j - \mu^i\|^2$$

Reformulăm folosind:

$$\min\{a_1,\cdots,a_N\} = \min_{s} \sum_{i} s_i a_i$$
$$\text{s.l.} \sum_{i} s_i = 1, \quad s \geq 0.$$

$$\min_{1\le i\le K} \|x^j - \mu^i\|^2 = \min_s \sum_i s_i \|x^j - \mu^i\|^2$$

$$\text{s.l.} \sum_i s_i = 1, \quad s \ge 0.$$

Problema $K-$*Means*:

$$\min_{\mu^i \in \mathbb{R}^n, s^j \in \mathbb{R}^K} \sum_{j=1}^{m} \sum_{i=1}^{K} s_i^j \|x^j - \mu^i\|^2$$

$$\text{s.l.} \sum_{i=1}^{K} s_i^j = 1, \quad s^j \ge 0.$$

Problema $K-Means$:

$$\min_{\mu^i \in \mathbb{R}^n, s^j \in \mathbb{R}^K} \sum_{j=1}^{m} \sum_{i=1}^{K} s_i^j \|x^j - \mu^i\|^2$$

$$\text{s.l.} \sum_{i=1}^{K} s_i^j = 1, \quad s^j \geq 0.$$

- problema este neconvexă în $(\mu, s)$
- subproblemele în $\mu$, respectiv $s$ sunt convexe
- intuitie: algoritm de minimizare alternativă

**Algoritm de minimizare alternativă**

Cât timp *criteriu oprire = FALS*:

1. $\mu^i_{k+1} = \arg\min \sum_{j=1}^{m} [s^j_k]_i \|x^j - \mu^i\|^2, \qquad \forall 1 \leq i \leq K$
2. $s^j_{k+1} = \arg\min \sum_{i=1}^{K} s^j_i \|x^j - \mu^i_{k+1}\|^2,$ s.l. $\sum_{i=1}^K s^j_i = 1, \;\; s^j \geq 0, \qquad \forall 1 \leq j \leq m$
3. $k := k + 1$

**Algoritm de minimizare alternativă**

Cât timp *criteriu oprire = FALS*:

1. $\mu^i_{k+1} = \frac{\sum\limits_{j=1}^{m} [s^j_k]_i \|x^j\|}{\sum\limits_{j=1}^{m} s^j_{i,k}}$
2. $[s^j_{k+1}]_i = \begin{cases} 1 & \text{dacă } i = \arg\min_i \|x^j - \mu^i\|^2 \\ 0 & \text{altfel} \end{cases}$
3. $k := k + 1$

[1 ] Christopher Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[2 ] A. Shapiro, *Introduction to Stochastic Programming*, SIAM.