Tehnici de Optimizare
Curs 3

Facultatea de Matematica si Informatica

Universitatea Bucuresti

- Department Informatica-

2021

Verificare convexitate

Sub non-diferentiabilitate: e.g. $\|Ax - b\|, \max\{f_1(x), \dots, f_m(x)\}, etc.$

Exemplu: $\left\||Ax - b|\right\|_\infty = \max\{|A_1x - b_1|, \dots, |A_mx - b_m|\}$

Verificam conditia de ordin 0:

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$$

Sub diferentiabilitate: e.g. $\frac{1}{2}x^T Hx + q^T x,\ \log(a^T x + b), etc.$

Verificam conditia de ordin II: $\nabla^2 f(x) \succcurlyeq 0, \quad \forall x$

Conditii de extrem

THEOREM 1 (Fermat). Let x^* be a minimum point of $f(x)$ on $\mathbf{R}^n$ and let $f(x)$ be differentiable at x^*. Then

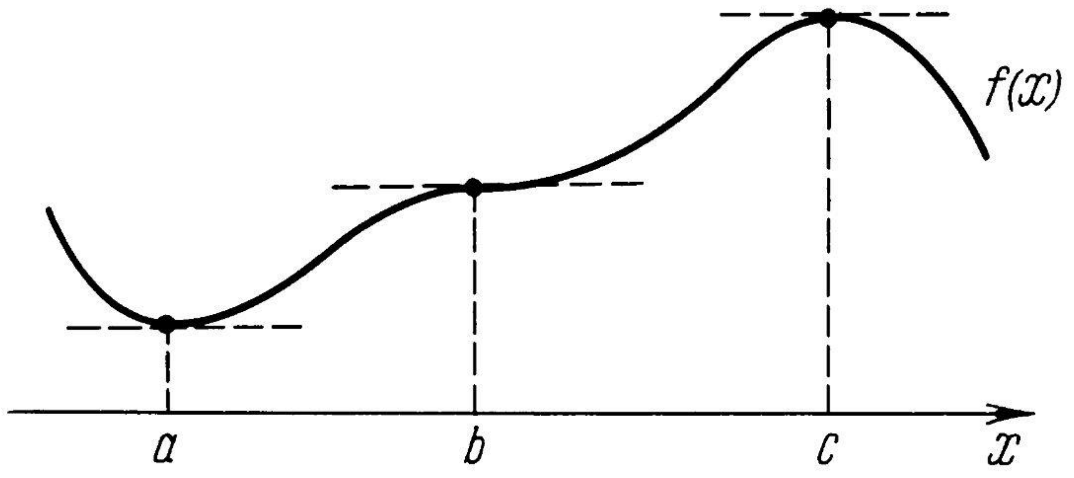
$$\nabla f(x^*) = 0 . \tag{1}$$

PROOF. Suppose $\nabla f(x^*) \neq 0$. Then

$$\begin{aligned} f(x^* - \tau \nabla f(x^*)) &= f(x^*) - \tau \|\nabla f(x^*)\|^2 + o(\tau \nabla f(x^*)) \\ &= f(x^*) - \tau(\|\nabla f(x^*)\|^2 + \tau^{-1} o(\tau)) < f(x^*) \end{aligned}$$

Conditii de extrem

- Demonstratia sugereaza de o metoda de constructie a unui punct pentru a descreste valoarea functiei:
$$x^+ = x - \alpha \nabla f(x)$$
- Pentru o alegere potrivita a lui α, avem $f(x^+) < f(x)$



Conditii ordin I suficiente

THEOREM 2. Let $f(x)$ be a convex function differentiable at a point x^* and let $\nabla f(x^*) = 0$. Then x^* is a global minimum point of $f(x)$ on $\mathbf{R}^n$.

PROOF. The proof follows immediately from formula (26) of Section 1.1 since $f(x) \geq f(x^*) + (\nabla f(x^*), x-x^*) = f(x^*)$ for any $x \in \mathbf{R}^n$. $\square$

Sub convexitate, orice punct de extrem este **punct de minim GLOBAL**!!!

Conditii necesare de ordin II

- Reamintim:

$$Diferentiabilitate\text{: } f(x+\tau \mathrm{d}) = f(x) + \tau \nabla f(x)^T d + o(\tau d)$$

$$Dublu\ dif: f(x+\tau d) = f(x) + \tau \nabla f(x)^T d + \tau^2 \mathrm{d}^{\mathrm{T}} \nabla^2 f(x) d + o(\tau^2 d)$$

Conditii necesare de ordin II

THEOREM 3. Let x^* be a minimum point of $f(x)$ on $\mathbf{R}^n$ and let $f(x)$ be twice differentiable at x^*. Then

$$\nabla^2 f(x^*) \geq 0 . \tag{2}$$

PROOF. By Theorem 1, $\nabla f(x^*) = 0$ and hence for an arbitrary y and a sufficiently small τ

$$f(x^*) \leq f(x^* + \tau y) = f(x^*) + \tau^2(\nabla^2 f(x^*)y, y)/2 + o(\tau^2) ,$$

$$(\nabla^2 f(x^*)y, y) \geq o(\tau^2)/\tau^2 .$$

Passing to the limit as $\tau \to 0$, we obtain $(\nabla^2 f(x^*)y, y) \geq 0$. Since y is arbitrary, $\nabla^2 f(x^*) \geq 0$. □

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

Exemplu

- $f: R^2 \rightarrow R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

$$\nabla f(x) = \begin{bmatrix} x_1^3 - x_1^2 \\ x_2^3 - x_2^2 \end{bmatrix}$$

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care sunt punctele stationare? Se poate gasi un minim local?

Puncte stationare: $\nabla f(x) = \begin{bmatrix} x_1^3 - x_1^2 \\ x_2^3 - x_2^2 \end{bmatrix}$=0

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

Puncte stationare: $\nabla f(x) = \begin{bmatrix} x_1^3 - x_1^2 \\ x_2^3 - x_2^3 \end{bmatrix} = 0 \Rightarrow x_1^*, x_2^* \in \{0,1\}$

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

Puncte stationare:$\left\{\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

Puncte stationare: $\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$

$$\nabla^2 f(x) = \begin{bmatrix} 3x_1^2 - 2x_1 & 0 \\ 0 & 3x_2^2 - 2x_2 \end{bmatrix}$$

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

Puncte stationare: $\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

Toate punctele stationare satisfac: $\nabla^2 f(x) = \begin{bmatrix} 3x_1^2 - 2x_1 & 0 \\ 0 & 3x_2^2 - 2x_2 \end{bmatrix} \succcurlyeq 0$

Conditii suficiente de ordin II

THEOREM 4. At a point x^*, let $f^*(x)$ be twice differentiable, let a first-order necessary condition hold (i.e., $\nabla f(x^*) = 0$) and let

$$\nabla^2 f(x^*) \; > \; 0 \, . \tag{3}$$

Then x^* is a local minimum point.

PROOF. Let y be any vector with unit norm. Then

$$\begin{aligned} f(x^* + \tau y) \; &= \; f(x^*) \; + \; \tau^2(\nabla^2 f(x^*)y, y)/2 \; + \; o(\tau^2 \|y\|^2) \\ &\geq \; f(x^*) \; + \; \tau^2 \ell/2 \; + \; o(\tau^2) \, , \end{aligned}$$

where $\ell > 0$ is the smallest eigenvalue of $\nabla^2 f(x^*)$ and the function $o(\tau^2)$ does not depend on y. Hence we can find a τ_0 such that for $0 \leq \tau \leq \tau_0$ we have $\tau^2 \ell/2 \geq o(\tau^2)$, i.e., $f(x^* + \tau y) \geq f(x^*)$. $\square$

Exemplu

- $f: R^2 \to R, f(x) = \frac{1}{4}(x_1^4 + x_2^4) - \frac{1}{3}(x_1^3 + x_2^3)$
- Care punctele stationare? Se poate gasi un minim local?

Puncte stationare:$\left\{\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$

Doar $[1\ 1]^T$ satisface cond. sufc.: $\nabla^2 f([1,1]) = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \succ 0$ $\Rightarrow$Minim local!

Proprietati ale punctelor de extrem

- Ce solutie are problema:

$$\min_{x} f(x) = \frac{1}{1+x^2}$$

Proprietati ale punctelor de extrem

- Ce solutie are problema:

$$\min_{x \geq \frac{1}{2}} f(x) = \frac{1}{1+x^2}$$

Convexitate: $\nabla^2 f(x) = \frac{8x^2-2}{(1+x^2)^2} \geq 0 \ (pe\ multimea\ fezabila)$

Proprietati ale punctelor de extrem

- Ce solutie are problema:

$$\min_{x \geq \frac{1}{2}} f(x) = \frac{1}{1+x^2}$$

Convexitate: $\nabla^2 f(x) = \frac{8x^2-2}{(1+x^2)^2} \geq 0 \ (pe\ multimea\ fezabila)$

Desi este convexa (pe mul. fez.), nu are puncte de minim!

Clasificare metode de optimizare

Informatia ce indica comportamentul unei functii $f \in \mathbb{R}^n \to \mathbb{R}$ intr-un punct $x \in \mathbb{R}^n$ se poate clasifica:

- Informatie de ordin 0: $f(x)$
- Informatie de ordin 1: $f(x), \nabla f(x)$
- Informatie de ordin 2: $f(x), \nabla f(x), \nabla^2 f(x)$
- ...

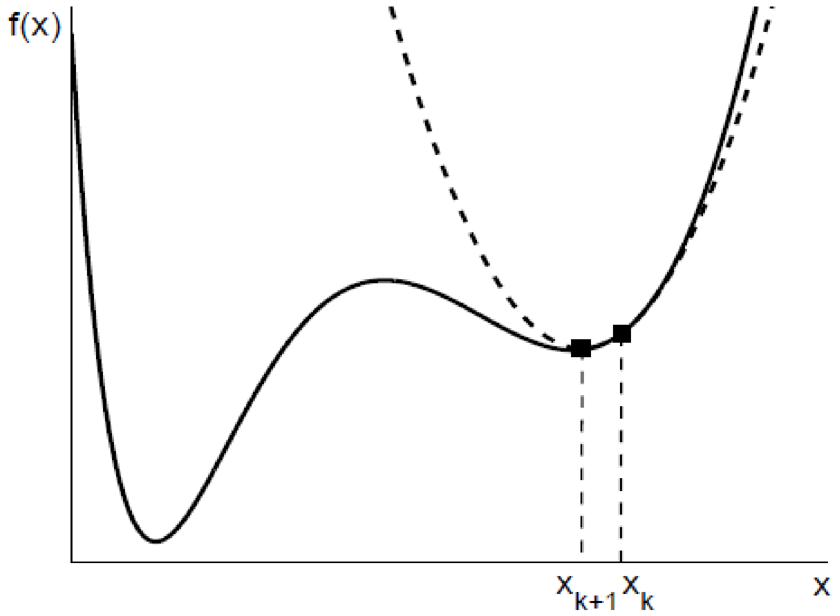
Fie algoritmul iterativ definit de $x_{k+1} = \mathcal{M}(x_k)$; in functie de ordinul informatiei utilizate in expresia lui $\mathcal{M}$:

- Metode de ordin 0: $f(x_k)$
- Metode de ordin 1: $f(x_k), \nabla f(x_k)$
- Metode de ordin 2: $f(x_k), \nabla f(x_k), \nabla^2 f(x_k)$

Metoda Gradient

- **Interpretare**: iteratia metodei gradient se obtine din minimizarea unei aproximari patratice a functiei obiectiv f

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\alpha_k} \|x - x_k\|^2$$



$$\nabla f(x^k) + \frac{1}{\alpha_k}(x - x^k) = 0$$
$$x - x^k = -\alpha_k \nabla f(x^k)$$
$$x^{k+1} = x^k - \alpha_k \nabla f(x^k)$$

Metoda Gradient

Cea mai "simpla" metoda de ordinul I: Metoda Gradient

- Prima aparitie in lucrarea [1] a lui Augustin-Louis Cauchy, 1847
- Cauchy rezolva un sistem neliniar de ecuatii cu 6 necunoscute, utilizand Metoda Gradient

[1] A. Cauchy. *Methode generale pour la resolution des systemes dequations simultanees.* C. R. Acad. Sci. Paris, 25:536-538, 1847

Imbunatatiri

Rata de convergenta slaba a metodei gradient reprezinta motivatia dezvoltarii de alte metode de ordin I cu performante superioare

- Metoda de Gradienti Conjugati - autori independenti Lanczos, Hestenes, Stiefel (1952)
  - QP convex solutia in n iteratii

- Metoda de Gradient Accelerat - dezvoltata de Yurii Nesterov (1983)

Continuitate Lipschitz

Fie o functie continuu diferentiabila f (i.e. $f \in \mathcal{C}^1$), atunci gradientul ∇f este **continuu Lipschitz** cu parametrul $L > 0$ daca:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in \operatorname{dom} f \tag{1}$$

Teorema 2: Relatia de Lipschitz (1) implica

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L}{2}\|x - y\|^2 \quad \forall x, y$$

Observatie: aceasta relatie este universal folosita in ratele de convergenta ale algoritmilor de ordinul I!

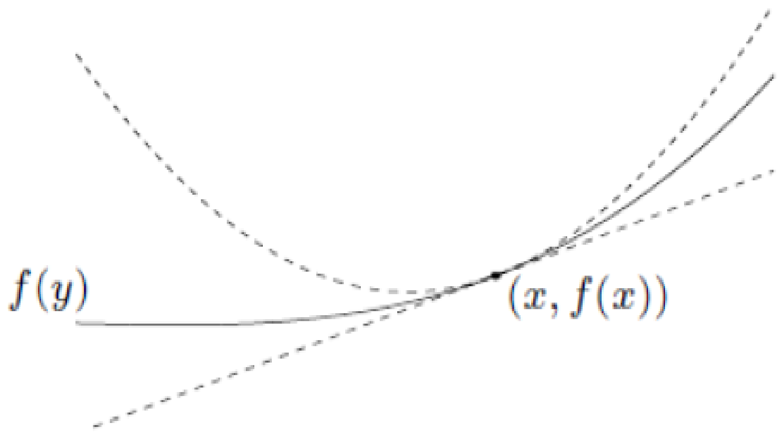
Teorema 3: In cazul functiilor de doua ori diferentiabile, relatia de Lipschitz (1) este echivalenta cu

$$\|\nabla^2 f(x)\| \leq L \quad \forall x \in \operatorname{dom} f$$

Continuitate Lipschitz

- Daca f are gradient Lipschitz atunci:

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|^2$$



Exemplu

Fie $f : \mathbb{R}^n \to \mathbb{R}$ o functie patratica, i.e.

$$f(x) = \frac{1}{2}x^T Qx + \langle q, x \rangle.$$

Observam expresia gradientului $\nabla f(x) = Qx + q$.

Aproximam constanta Lipschitz a functiei f:

$$\|Qx + q - Qy - q\| = \|Q(x - y)\| \leq \|Q\|\|x - y\| = L\|x - y\|$$

In concluzie, pentru functiile patratice constanta Lipschitz este:

$$L = \|Q\| = \lambda_{\max}(Q)$$

Exemplu

Fie $f : \mathbb{R}^n \to \mathbb{R}$ definita de

$$f(x) = \log\left(1 + e^{a^T x}\right).$$

Observam expresia gradientului si a matricii Hessiene

$$\nabla f(x) = \frac{e^{a^T x}}{1 + e^{a^T x}} a \qquad \nabla^2 f(x) = \frac{e^{a^T x}}{(1 + e^{a^T x})^2} a a^T$$

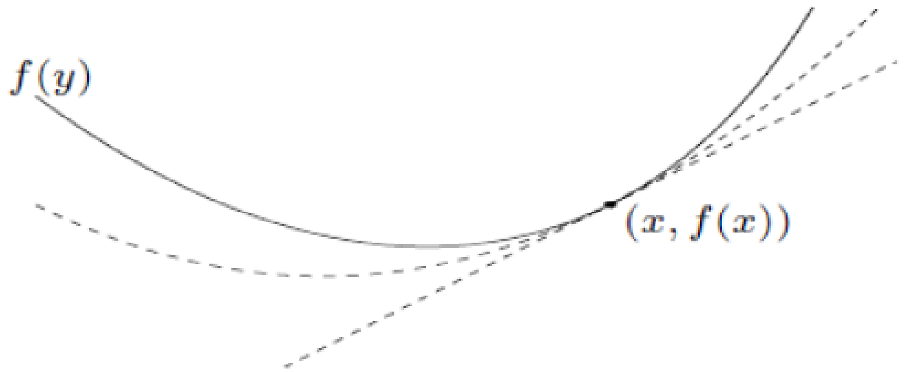
Pentru orice constanta pozitiva $c > 0$ avem $\frac{c}{(1+c)^2} \leq \frac{1}{4}$, deci

$$\|\nabla^2 f(x)\| = \frac{e^{a^T x}}{(1 + e^{a^T x})^2} \|a a^T\| \leq \frac{\|a\|^2}{4} = L$$

Convexitate tare

- Daca f este tare convexa atunci:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\sigma}{2} \|y - x\|^2$$



Garantii de convergenta

- Fie f diferentiabila cu ∇f *Lipschitz continuu* (constanta Lipschitz $L > 0$)
- Alegem lungimea pasului $\alpha_k = \frac{1}{L}$

Atunci rata de convergenta globala a sirului x_k generat de metoda gradient este subliniara, data de:

$$\min_{0 \leq i \leq k} \|\nabla f(x_i)\| \leq \frac{1}{\sqrt{k}} \sqrt{2L(f(x_0) - f^*)}$$

Garantii de convergenta

Sub presupunerea ca ∇f Lipschitz continuu avem:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 \quad x, y \in \mathrm{dom} f.$$

Considerand $x = x_k, y = x_{k+1} = x_k - (1/L)\nabla f(x_k)$ avem:

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Insumam dupa $i = 0, \ldots k - 1$ si rezulta

$$\frac{1}{2L}\sum_{i=0}^{k-1}\|\nabla f(x_i)\|^2 \leq f(x_0) - f(x_k) \leq f(x_0) - f^*$$

In concluzie, observam

$$k \min_{0 \leq i \leq n}\|\nabla f(x_i)\|^2 \leq \sum_{i=0}^{k-1}\|\nabla f(x_i)\|^2 \leq 2L(f(x_0) - f^*)$$

Garantii de convergenta a sirului

- Fie f diferentiabila cu ∇f *Lipschitz continuu* (constanta Lipschitz $L > 0$)
- Exista un punct de minim local x^*, astfel incat Hessiana in acest punct satisface

$$\sigma I_n \preceq \nabla^2 f(x^*) \preceq L I_n$$

- Punctul initial x_0 al iteratiei metodei gradient cu pas $\alpha_k = \frac{2}{\sigma + L}$ este suficient de aproape de punctul de minim, i.e.

$$\|x_0 - x^*\| \leq \frac{2\sigma}{L}$$

Atunci rata de convergenta locala a sirului x_k generat de metoda gradient este liniara (i.e de ordinul $\mathcal{O}\left(\log(\frac{1}{\epsilon})\right)$), data de:

$$\|x_k - x^*\| \leq \beta \left(1 - \frac{2\sigma}{L + 3\sigma}\right)^k \qquad \text{cu } \beta > 0$$

Garantii de convergenta a sirului

- Fie f diferentiabila cu ∇f *Lipschitz continuu* (constanta Lipschitz $L > 0$)
- Exista un punct de minim local x^*, astfel incat Hessiana in acest punct satisface

$$\sigma I_n \preceq \nabla^2 f(x^*) \preceq L I_n$$

- Punctul initial x_0 al iteratiei metodei gradient cu pas $\alpha_k = \frac{2}{\sigma + L}$ este suficient de aproape de punctul de minim, i.e.

$$\|x_0 - x^*\| \leq \frac{2\sigma}{L}$$

Atunci rata de convergenta locala a sirului x_k generat de metoda gradient este liniara (i.e de ordinul $\mathcal{O}\left(\log(\frac{1}{\epsilon})\right)$), data de:

$$\|x_k - x^*\| \leq \beta \left(1 - \frac{2\sigma}{L + 3\sigma}\right)^k \qquad \text{cu } \beta > 0$$

Garantii de convergenta a sirului

- Fie f functie convexa, diferentiabila cu ∇f Lipschitz continuu (constanta Lipschitz $L > 0$). Daca alegem lungimea pasului constanta $\alpha_k = \frac{1}{L}$, atunci rata de convergenta globala a sirului x_k generat de metoda gradient este subliniara, data de:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|^2}{2k}$$

- Daca in plus functia este tare convexa cu constanta $\sigma > 0$, atunci rata de convergenta globala a sirului x_k generat de metoda gradient cu pas $\alpha_k = \frac{1}{L}$ este liniara, data de:

$$\|x_k - x^*\|^2 \leq \left(\frac{L - \sigma}{L + \sigma}\right)^k \|x_0 - x^*\|^2$$

Garantii de convergenta a sirului

$\boldsymbol{\mu = \frac{L}{\sigma}}$ **(numarul de conditionare al problemei) determina viteza algoritmilor de ordin I**

Cand μ este mare, atunci MG necesita un numar mare de iteratii!
Cand μ este mic, atunci MG necesita un numar mic de iteratii!