

*Optimizarea matematica numita adesea
"Science of the Better"*

(UNLP)

Problema de optimizare neliniara fara constrangeri

$$(UNLP) : \quad f^* = \min_{x \in \mathbb{R}^n} f(x)$$

Condițiile de ordinul I: $\nabla f(x) = 0$ (n ecuatii cu n necunoscute)

Exemplu (QP):

$$\min_{x \in \mathbb{R}^n} 0.5x^T Qx - q^T x,$$

- ▶ Q matrice inversabila $\implies \nabla f(x) = Qx - q = 0$ solutie unica
 $x^* = Q^{-1}q$
- ▶ valoarea optima $f^* = -0.5q^T Q^{-1}q$.

Metode numerice de optimizare

$$(UNLP) : \quad f^* = \min_{x \in \mathbb{R}^n} f(x)$$

Sistemul de n ecuatii cu n necunoscute $\nabla f(x) = 0$ nu poate fi rezolvat analitic! \implies *metode iterative de rezolvare*

Algorithm iterativ: genereaza un sir $(x_k)_{k \geq 0}$; fiecare punct calculat pe baza punctelor anterioare:

$$x_{k+1} = \mathcal{M}(x_0, \dots, x_k)$$

- ▶ *algorithm convergent global:* pentru orice x_0 sirul x_k converge la o solutie x^*
- ▶ *algorithm convergent local:* pentru orice x_0 apropiat de o solutie x^* sirul x_k converge la x^*

Metode numerice de optimizare

$$(UNLP): \quad f^* = \min_{x \in \mathbb{R}^n} f(x)$$

Algorithm iterativ: genereaza un sir de puncte $(x_k)_{k \geq 0}$; fiecare punct calculat pe baza punctelor anterioare:

$$x_{k+1} = \mathcal{M}(x_0, \dots, x_k)$$

Algoritmul numeric colecteaza informatie printr-un *oracol* si manipuleaza raspunsurile oracolului

- ▶ *oracol de ordin zero* \mathcal{O}_0 : furnizeaza informatie bazata pe evaluarea functiei obiectiv $f(x)$
- ▶ *oracol de ordin intai* \mathcal{O}_1 : furnizeaza informatie bazata pe evaluarea functiei obiectiv $f(x)$ si gradientului $\nabla f(x)$
- ▶ *oracol de ordin doi* \mathcal{O}_2 : furnizeaza informatie bazata pe evaluarea functiei obiectiv $f(x)$, gradient $\nabla f(x)$ si Hessiana $\nabla^2 f(x)$

Metode numerice de optimizare

Algoritm de optimizare iterativ gaseste o solutie aproximativa cu o acuratete prestabilita ϵ

acuratetea ϵ reprezinta si *criteriul de oprire* al algoritmului

- ▶ criteriul de oprire 1: $\|\nabla f(x_k)\| \leq \epsilon$
- ▶ criteriul de oprire 2: $|f(x_k) - f^*| \leq \epsilon$
- ▶ criteriul de oprire 3: $\|x_{k+1} - x_k\| \leq \epsilon$

Metoda generica de optimizare

1. se da: punct initial x_0 , acuratete ϵ si contor $k = 0$
2. \mathcal{I}_k informatia acumulata de oracol la pasul k ; se apeleaza oracolul \mathcal{O} in x_k ; se actualizeaza informatia $\mathcal{I}_{k+1} = \mathcal{I}_k \cup \mathcal{O}(x_k)$
3. se calculeaza $x_{k+1} = \mathcal{M}(\mathcal{I}_{k+1})$

Exemplu (metoda gradient): $\mathcal{O}(x_k) = \nabla f(x_k) \implies$

$x_{k+1} = x_k - \alpha \nabla f(x_k)$, cu scalar $\alpha > 0$ ales la pasul initial $k = 0$

Metode numerice de optimizare

- **Complexitatea analitica a algoritmului de optimizare**

iterativ: numarul total de apeluri ale oracolului pentru gasirea unei ϵ -solutii

- **Rata de convergenta:** viteza cu care sirul x_k converge la solutia x^* ; ordinul de convergenta q definit ca:

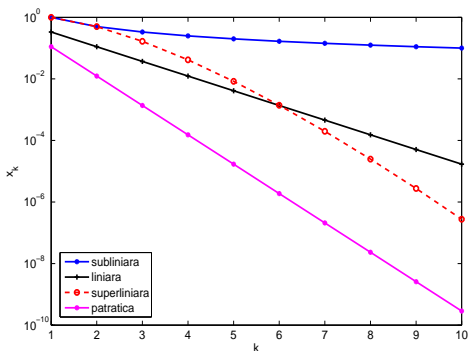
$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^q} < \infty \iff \|x_{k+1} - x^*\| \approx \beta \|x_k - x^*\|^q$$

- ▶ *convergenta subliniara:* $\|x_k - x^*\| \leq \frac{\beta}{k^q}$
- ▶ *convergenta liniara* ($q = 1$): $\|x_{k+1} - x^*\| \leq \beta \|x_k - x^*\|$, cu $0 < \beta < 1$
- ▶ *convergenta superliniara* ($q = 1$):
 $\|x_{k+1} - x^*\| \leq \beta_k \|x_k - x^*\|$, cu $\beta_k \rightarrow 0$
- ▶ *convergenta patratica* ($q = 2$): $\|x_{k+1} - x^*\| \leq \beta \|x_k - x^*\|^2$, cu $\beta > 0$

Rata de convergenta

$\|x_{k+1} - x^*\| \approx \beta \|x_k - x^*\|^q \iff$ distanta se reduce cu q zecimale

- ▶ *convergenta subliniara*: $x_k = \frac{1}{k} \Rightarrow (1; 0.5; 0.33; 0.25; 0.2)$
- ▶ *conv. liniara*: $x_k = \frac{1}{3^k} \Rightarrow (0.33; 0.11; 0.03; 0.01; 0.004)$
- ▶ *conv. superliniara*: $x_k = \frac{1}{3^{k!}} \Rightarrow (1; 0.16; 0.05; 0.01; 0.002)$
- ▶ *conv. patratica*: $x_k = \frac{1}{3^{2^k}} \Rightarrow (0.1; 0.01; 0.001; 0.0002; 0.00001)$



Convergenta metodelor de optimizare

Definitie. Se da: spatiul metric (X, ρ) , submultimea $S \subseteq X$ si metoda \mathcal{M} de tip aplicatie punct-multime $(\mathcal{M} : X \rightarrow 2^X)$. Definim functia descrescatoare $\phi : X \rightarrow \mathbb{R}$ pentru perechea (S, \mathcal{M}) :

- ▶ $\forall x \in S$ si $y \in \mathcal{M}(x)$ avem $\phi(y) \leq \phi(x)$
- ▶ $\forall x \notin S$ si $y \in \mathcal{M}(x)$ avem $\phi(y) < \phi(x)$

Exemplu: pentru problema $\min_x f(x)$, definim

$S = \{x : \nabla f(x) = 0\}$ (multimea punctelor stationare) si $\phi = f$ (de obicei o metoda de optimizare alege x_{k+1} a.i. $f(x_{k+1}) \leq f(x_k)$)

Definitie. O aplicatie punct-multime $\mathcal{M} : X \rightarrow 2^X$ este inchisa in x_0 daca pentru orice $x_k \rightarrow x_0$ si $y_k \rightarrow y_0$ cu $y_k \in \mathcal{M}(x_k)$ avem $y_0 \in \mathcal{M}(x_0)$. \mathcal{M} este inchisa daca este inchisa in toate punctele din X .

Exemplu: aplicatia punct-multime $x_{k+1} \in [-|x_k|, |x_k|]$ este inchisa

Convergenta metodelor de optimizare

Theorema de convergenta generala.

Fir sirul $x_{k+1} \in \mathcal{M}(x_k)$ iar S multimea solutiilor, satisfacand:

- ▶ x_k se afla intr-o multime compacta
- ▶ \mathcal{M} este aplicatie punct-multime inchisa
- ▶ exista o functie continua descrescatoare ϕ pentru (\mathcal{M}, S)

Atunci punctele limita ale sirului x_k apartin multimii solutiilor S .

Metode de descrestere

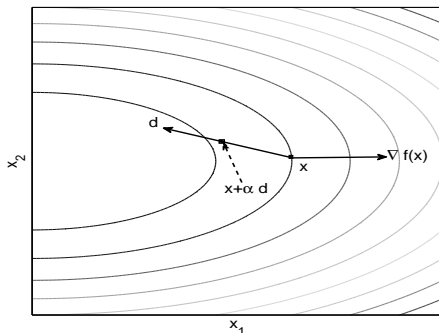
$$x_{k+1} = x_k + \alpha_k d_k,$$

unde d_k este directie de descrestere pentru f in x_k , i.e.

$$\nabla f(x_k)^T d_k < 0$$

pentru $\alpha_k \Downarrow$ suficient de mic

$$f(x_{k+1}) \leq f(x_k)$$



Metode de descrestere

$$x_{k+1} = x_k + \alpha_k d_k \implies \nabla f(x_k)^T d_k < 0$$

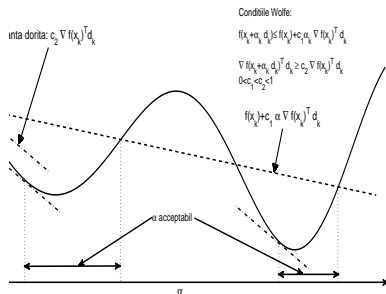
Strategii de alegerea a pasului α_k

- ▶ metoda ideala: $\alpha_k = \arg \min_{\alpha \geq 0} f(x_k + \alpha d_k)$
- ▶ metoda Wolfe:

$$(W1): f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T d_k \quad \text{cu } c_1 \in (0, 1)$$

$$(W2): \nabla f(x_k + \alpha_k d_k)^T d_k \geq c_2 \nabla f(x_k)^T d_k \quad \text{cu } c_1 < c_2 < 1$$

- ▶ metoda backtracking: alegem $\alpha > 0$ si $\rho, c_1 \in (0, 1)$
cat timp $f(x_k + \alpha d_k) > f(x_k) + c_1 \alpha \nabla f(x_k)^T d_k \Rightarrow \alpha = \rho \alpha$



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:

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

Lema 1: 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 II!

Lema 2: In cazult functiilor de doua ori diferentiabile (i.e. $f \in \mathcal{C}^2$), relatia de Lipschitz (1) este echivalenta cu

$$\nabla^2 f(x) \preceq LI_n \iff \|\nabla^2 f(x)\| \leq L \quad \forall x \in \text{dom} f$$

Continuitate Lipschitz

Demonstratie Lema 1: Din teorema valorii medii, pentru orice x, y avem:

$$\begin{aligned}f(y) &= f(x) + \int_0^1 \langle \nabla f(x + \tau(y - x)), y - x \rangle d\tau \\&= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle d\tau\end{aligned}$$

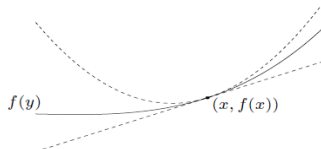
Evaluand modulul aproximarii liniare, avem

$$\begin{aligned}&|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \\&= \left| \int_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle d\tau \right| \\&\leq \int_0^1 |\langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle| d\tau \\&\leq \int_0^1 \tau L \|x - y\|^2 d\tau = \frac{L}{2} \|x - y\|^2\end{aligned}$$

Continuitate Lipschitz - convexitate tare

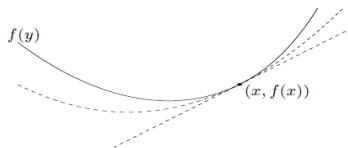
- Dacă f are gradient Lipschitz atunci:

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



- Dacă f este tare convexa atunci:

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



- Dacă f tare convexa și gradient Lipschitz atunci co-coercivitate:

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \frac{\sigma L}{\sigma + L} \|x - y\|^2 + \frac{1}{\sigma + L} \|\nabla f(y) - \nabla f(x)\|^2$$

Continuitate Lipschitz - Exemplu 1

Fie $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 patratiche constanta Lipschitz este:

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

Exemplu: daca $Q = \text{diag}([q_1 \ q_2])$, cu $q_1 > q_2 > 0$, atunci constanta Lipschitz $L = q_1$ si constanta de convexitate tare $\sigma = q_2$.

Continuitate Lipschitz - Exemplu 2

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

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

Observam expresia gradientului si a matricii Hessiene

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

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

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

Convergenta metodelor de descrestere

Teorema de convergenta globala. Metoda de descrestere, cu pasul satisfacand conditiile Wolfe, aplicata problemei (UNLP) cu functia obiectiv marginita inferior si avand gradientul Lipschitz are urmatoarea convergenta globala:

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < \infty,$$

unde θ_k este unghiul facut de directia d_k cu gradientul $\nabla f(x_k)$.

Demonstratie:

Din conditia Wolfe (W2) avem:

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k \geq (c_2 - 1) \nabla f(x_k)^T d_k.$$

Aplicand inegalitatea Cauchy-Schwartz obtinem:

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \cdot \|d_k\| \geq (c_2 - 1) \nabla f(x_k)^T d_k.$$

Din proprietatea de Lipschitz gradient avem:

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq L \|x_{k+1} - x_k\| = L \alpha_k \|d_k\|$$

Convergenta metodelor de descrestere

Demonstratie (continua):

Inlocuim in relatia anterioara ultima inegalitate:

$$L\alpha_k \|d_k\|^2 \geq (c_2 - 1) \nabla f(x_k)^T d_k,$$

i.e.

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{\nabla f(x_k)^T d_k}{\|d_k\|^2}.$$

Din conditia Wolfe (W1) avem:

$$f(x_{k+1}) \leq f(x_k) + c_1 \frac{(\nabla f(x_k)^T d_k)^2}{\|d_k\|^2} \frac{c_2 - 1}{L}$$

ceea ce conduce la

$$f(x_{k+1}) \leq f(x_k) - c \cos^2 \theta_k \|\nabla f(x_k)\|^2,$$

unde $c = c_1(1 - c_2)/L$. Insumand de la $k = 0$ la $N - 1$ obtinem

$$f(x_N) \leq f(x_0) - c \sum_{j=0}^{N-1} \cos^2 \theta_j \|\nabla f(x_j)\|^2.$$

In final, tinem cont ca f este marginita inferior si luam $N \rightarrow \infty$.