Tehnici de Optimizare
Condiţii de extrem. Algoritmi de optimizare

Fie funcţia $f(x) = x^4 + 7x^3 + 5x^2 - 17x + 3$.

- Care sunt punctele de extrem ale functiei?

$$f'(x) = 4x^3 + 21x^2 + 10x - 17 = 0$$

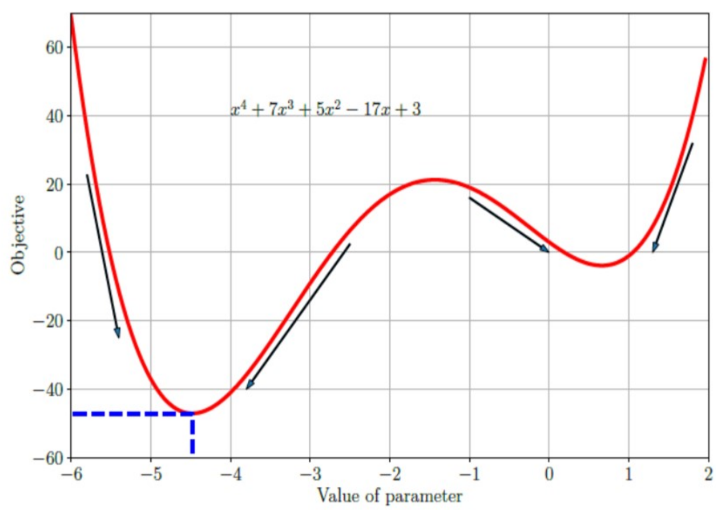
$(x^*)^1 = 0.66238, (x^*)^2 = -1.43211, (x^*)^3 = -4.48027.$

- Ce natura au aceste puncte?

  $f'(x)$ creste cand $f''(x) > 0$

$$f''((x^*)^1) = 43.0849271728 \quad -25.537351374800004 \;\; 62.702491274799996$$

- Intuiti vreo proprietate a punctelor de extrem pe $\mathbb{R}^n$?



1 Condiţii de ordin I

Theorem 1 (Fermat)**.** *Fie x^* un punct de minim al funcţiei f (pe $\mathbb{R}^n$). Considerând că f este diferenţiabilă în x^* atunci:*

$$\nabla f(x^*) = 0.$$

$$f(x) = \frac{1}{2}x^T Hx + q^T x \Rightarrow \underbrace{Hx + q}_{\nabla f(x)} = 0$$

Sub convexitate condiţiile de ordin I devin suficiente pentru gasirea extremelor.

Theorem 2. *Fie f o funcţie convexă şi diferenţiabilă şi un punct x^* ce satisface $\nabla f(x^*) = 0$. Atunci x^* este punct de minim global al funcţiei f (pe $\mathbb{R}^n$).*

2 Condiţii de ordin II

Condiţiile necesare de ordin I definesc *punctele staţionare* ale unei funcţii. Se pot investiga mai departe condiţiile de extrem, folosind derivate superioare.

Theorem 3 (Condiţii necesare de ordin II)**.** *Fie x^* un punct de minim al funcţiei f (pe $\mathbb{R}^n$). Considerând că f este de două ori diferenţiabilă în x^* atunci:*

$$\nabla^2 f(x^*) \succeq 0.$$

Theorem 4 (Condiţii suficiente de ordin II)**.** *Fie f dublu diferenţiabilă în x^*. Considerăm că x^* satisface condiţiile de ordin I ($\nabla f(x^*) = 0$) şi:*

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

Atunci x^* este un minim local.

$$\nabla^2 f(x^*) \succ qI_n \;\; (q > 0).$$

Observaţi ca dacă $\nabla f(x^*) = 0$ şi $\nabla^2 f(x^*) \succeq 0$, dar condiţiile suficiente de ordin II nu au loc, atunci nu este necesar ca x^* să fie un minim. Luaţi ca exemplu $f(x) = x^3, x \in \mathbb{R}$.

$$f'(x) = 3x^2 = 0 \Rightarrow x^* = 0$$
$$f''(0) = 0$$

Numar de conditionare al unui punct de minim.

3 Algoritmi de optimizare

Fie $f : \mathbb{R} \to \mathbb{R}$ diferenţiabilă, i.e.

$$\begin{aligned} f(x + \alpha d) &= f(x) + \alpha f'(x)d + o(\alpha d) \\ &= f(x) + \alpha \left(f'(x)d + \underbrace{\frac{o(\alpha d)}{\alpha}}_{\alpha \rightrightarrows 0\, 0} \right) \\ [alegem\ \ d = -f'(x)] &= f(x) + \alpha \left(-(f'(x))^2 + \underbrace{\frac{o(\alpha(-f'(x)))}{\alpha}}_{\alpha \rightrightarrows 0\, 0} \right) < f(x)\ [folosind\ \tilde{x} = x - \alpha f'(x), pentru\ \alpha\ mic] \end{aligned}$$

Intuiţi o valoare a direcţiei d care scade al doilea termen?

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

Dar pentru funcţiile cu gradient Lipschitz? Reamintim: *o funcţie are gradient Lipschitz dacă:*

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

sau echivalent

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

3.1 Metoda Gradient

In principiu, metoda gradient reprezintă un algoritm de ordin I care generează un şir de puncte (vectori) $x^1, x^2, ...$, pornind dintr-un punct iniţial ales x^0.

Metoda Gradient(x^0, ϵ)

Initializeaza $k = 0$.
Cat timp *criteriu_stop* :
1. Calculeaza $\nabla f(x^k)$
2. Actualizeaza: $x^{k+1} = x^k - \alpha_k \nabla f(x^k)$
3. Set $k := k + 1$.

Criteriu de stop: $\|\nabla f(x^k)\| \leq \epsilon$

La primul pas 1. (din algoritmul de mai sus), presupunem că un oracol de ordin I returnează gradientul ∇f evaluat în punctul curent x^k. Un aspect important al metodei este alegerea pasului $\alpha_k > 0$ dintre urmatoarele optiuni:

1. **Constant**: $\alpha_k = \alpha > 0$. In cazul in care f are gradient Lipschitz atunci pentru $\alpha \in \left(0, \frac{2}{L_f}\right)$:

$$f(x^{k+1}) \leq f(x^k) - \alpha \left(1 - \frac{L_f \alpha}{2}\right) \|\nabla f(x^k)\|^2.$$

Observam ca cea mai mare descrestere are loc pentru pasul constant optim $\alpha = \frac{1}{L_f}$, anume

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L_f}\|\nabla f(x^k)\|^2.$$

2. **Ideal**: $\alpha_k = \arg\min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k))$

3. **Backtracking**: alege $c > 0$, ajustează pasul α_k astfel incat sa aiba loc relaţia de descrestere:

$$f(x^k - \alpha_k \nabla f(x^k)) \leq f(x^k) - c\alpha_k \|\nabla f(x^k)\|^2. \quad (1)$$

procedura presupune alegerea unui parametru $\rho \in (0, 1]$ si actualizarea pasului, dupa cum urmeaza:

(i) Alegem $c, \rho \in (0,1), \alpha_{k,0} > 0$

(ii) Cat timp $\alpha_{k,t}$ nu satisface (2) iteram: $\alpha_{k,t+1} = \rho\alpha_{k,t}; \quad t = t+1$

Ce spune teoria?

Cum variază convergenţa cu:

- *numărul de conditionare* al punctului de minim
- pasul

3.1.1 Probleme rezolvate

1. Fie funcţia $f : \mathbb{R}^2 \to \mathbb{R}, f(x) = \frac{1}{2}\|Ax - b\|_2^2 + 2x^T x$, unde $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Sa se aducă la forma standard QP şi să se calculeze prima iteraţie a metodei gradient.
Rezolvare.

$$\begin{aligned} f(x) &= \frac{1}{2}\|Ax - b\|_2^2 + 2x^T x \\ &= \frac{1}{2}(Ax - b)^T(Ax - b) + 2x^T x \\ &= \frac{1}{2}x^T A^T Ax - b^T Ax + 2x^T I_n x + \frac{1}{2}b^T b \\ &= \frac{1}{2}x^T(A^T A + 4I_n)x - (A^T b)^T x + \frac{1}{2}b^T b \end{aligned}$$

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$
$$\|\nabla^2 f(x)\| \leq L$$
$$\|A^T A + 4I_n\| = \|\begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix}\| \leq L = 6$$

Plecam $x^0 = [-1\ 1]$, calculam x^1 al MG cu $\alpha = \frac{1}{L} = \frac{1}{6}$:

$$x^1 = x^0 - \alpha\nabla f(x^0) = x^0 - \alpha[(A^T A + 4I_n)x^0 - A^T b] = \begin{bmatrix} 1 - \frac{1}{2} \\ -1 + \frac{5}{6} \end{bmatrix}$$

$$f(x^0) = 1/2 + 4, f(x^1) = 1/2(1/9 + 4/9) + (1/2) * 11/9$$

3.2 Metoda Newton

Metoda Newton(x^0, ϵ)
Initializeaza $k = 0$. Cat timp *criteriu_stop* : 1. Calculeaza $\nabla f(x^k), \nabla^2 f(x^k)$ 2. Actualizeaza: $x^{k+1} = x^k - \alpha_k[\nabla^2 f(x^k)]^{-1}\nabla f(x^k)$ 3. Set $k := k + 1$.

Iteratia:

$$x^{k+1} = x^k - \alpha_k[\nabla^2 f(x^k)]^{-1}\nabla f(x^k)$$

de fapt se obtine din rezolvarea sistemului $\nabla^2 f(x^k)d = \nabla f(x^k)$ si apoi $x^{k+1} = x^k - \alpha_k d$

Alegeri pas:

- constant: alegerea standard $\alpha = 1$
- ideal: $\alpha_k = \arg\min_{\alpha>0} f(x^k - \alpha[\nabla^2 f(x^k)]^{-1}\nabla f(x^k))$
- Backtracking: alege $c > 0$, ajustează pasul α_k astfel incat sa aiba loc relaţia de descrestere:

  $$f(x^k - \alpha_k \nabla f(x^k)) \leq f(x^k) - c\alpha_k \nabla f(x^k)^T [\nabla^2 f(x^k)]^{-1}\nabla f(x^k)), \tag{2}$$

  asemănător cu cazul metodei gradient.

Ce spune teoria?

3.2.1 Probleme rezolvate

1. Fie funcţia $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x) = \frac{1}{4}\left(x_1^4 + x_2^4\right) - \frac{1}{3}\left(x_1^3 + x_2^3\right)$$

a) Determinaţi un minim local al funcţiei f.
Rezolvare.

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

Puncte stationare:

$$\nabla f(x) = \begin{bmatrix} x_1^3 - x_1^2 \\ x_2^3 - x_2^2 \end{bmatrix} = 0 \Leftrightarrow x_1, x_2 \in \{0, 1\}$$

$$\nabla^2 f(x) \in \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\},$$

$x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ minim local!

b) Calculaţi prima iteraţie a metodei Newton pornind din $x^0 = [-1\ 1]^T$ şi considerând cazurile: $\alpha_0 = 1$ şi α_0 ideal.