Ingineria Sistemelor cu Inteligenta Artificiala

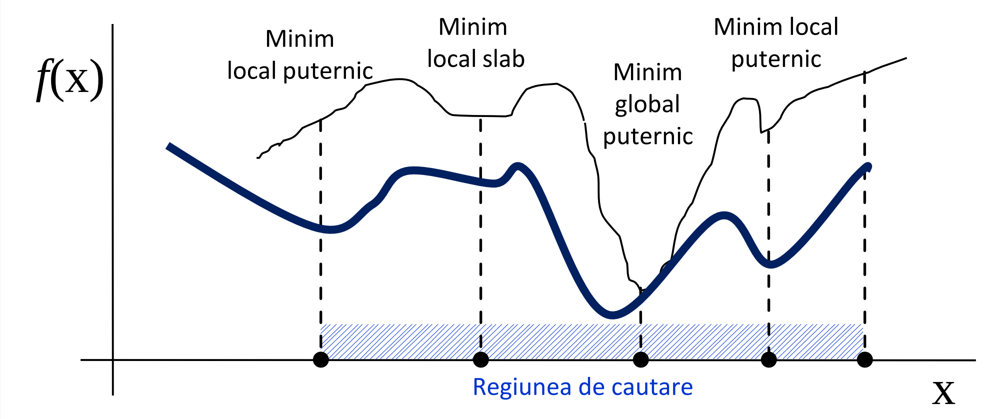
Corneliu Florea.

Tehnici de OPTIMIZARE

Gradient Descent

Preliminarii

- Curs (disciplina) in anul 3 dedicat la CTI



Problema:

Fiind data o functie f(x)- cat este x astfel incat functia sa fie minima

Optim: minim sau maxim

x poate fi scalar sau vectorial

f poate fi scalara sau vectoriala

- Care dintre minime va fi gasit depinde de punctul de pornire
- Situatiile ilustrate se regasesc toate in practica

Categorii pentru probleme de decizie

Categoria 1:

- Setul tuturor alternativelor posibile este discret (cu un numar redus de valori)
  - Exemplu: “Un student poate alege intre patru cursuri , toate cu referinte bune, dar trebuie sa opteze pentru unul singur.”
- Solutia: metode de evaluare (scoring methods)

Categoria 2:

- Numarul de alternative posibile este foarte mare, daca nu chiar infinit; decizia trebuie sa satisfaca si o serie de constangeri suplimentare
- Solutia: metode de optimizare (cu constrangeri)

Exemplu de metode notare (scoring)

Alfredo trebuie sa aleaga intre patru cursuri

Caracteristica	Nota pt cursul				Tip scor	Domeniu	Pondere
	C1	C2	C3	C4			
Utilitate	7	8	9	4	Pozitiv	0-10	7
Claritate predare	5	7	9	5	Pozitiv	0-10	3
Usurinta examen	5	7	9	1	Negativ	0-10	10
Merg prieteni	5	4	2	8	Pozitiv	0-10	7

Categoria 2 Probleme de optimizare

1. Se construieste o formulare clara a problemei se culeg toate datele relevante.
    1. Factori necontrolabili (variabile aleatoare)
    2. Data controlabile (variabile deterministe)

1. Se construieste un model matematic (pb de optimizare) .
    - Se formuleaza (defineste) functia obiectiv si constrangerile

1. Se rezolva modelul
    - Se aplica cel mai potrivit algorim pentru rezolvarea problemei

1. Implementare

Definirea problemei

Avem functia obiectiv de optimizat

$$f(\mathbf{x}) : \mathbb{R}^N \longrightarrow \mathbb{R}$$

Scopul este sa aflam variabila determinista $\mathbf{x}$ care minimizeaza functia (i.e. pentru care f atinge cea mai mica valoare)

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x})$$

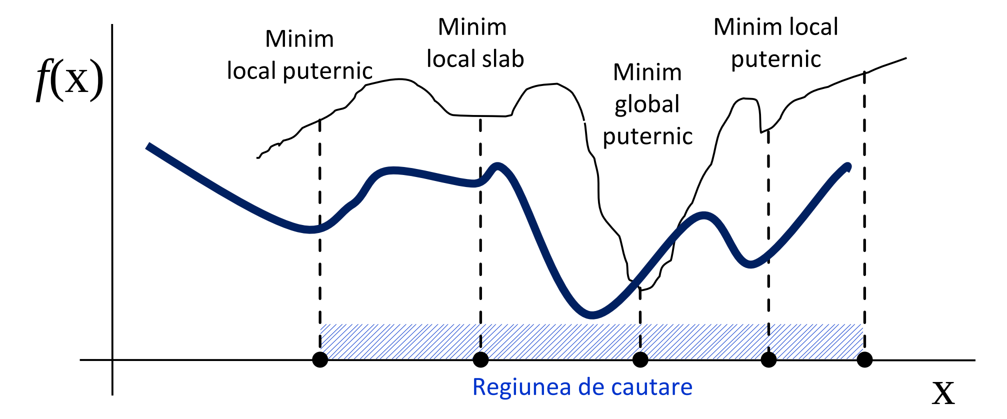
Minizarea necesita constrangerile:

- De tip egalitate: $c_i(\mathbf{x}) = 0$
- De tip inegalitate: $c_j(\mathbf{x}) \geq 0$

Problema: Vrem sa alegem drept presedinte cel mai bun politician roman.
Constrangere 1: in viata
Constrangere 2: Cu studii de cel putin 5 ani

Maximizarea (functie de tip profit) este echivalenta cu a cauta minimul lui $-f(\mathbf{x})$

Tipuri de minime

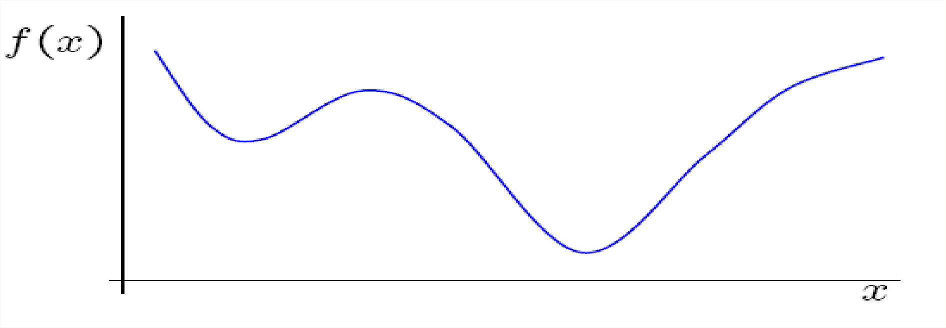


- Care dintre minime va fi gasit depinde de punctul de pornire
- Situatiile ilustrate se regasesc toate in practica

Optimizare univariata fara constrangeri

Vom presupune ca pornim aproape de minimul global

$$\min_x f(x)$$

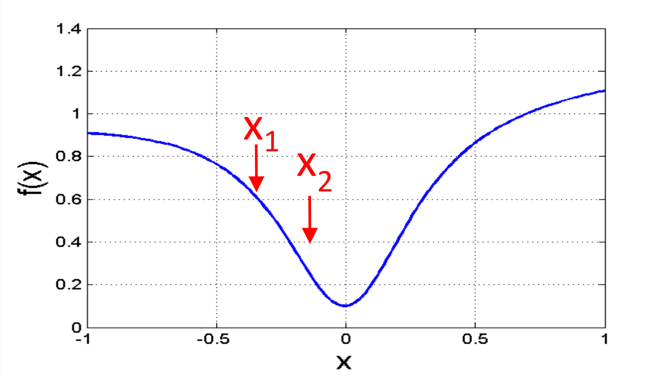


Pentru a determina pozitia minimului?

- Metode de cautare (Dichotomous, Fibonacci, Sectiunea de aur)
- Metode de aproximare
    1. Interpolare polinomiala
    2. Newton de tip Newton
- Combinatii

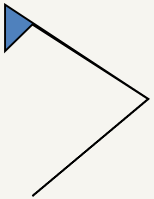
Metode de optimizare: intuitie

- Se porneste de la un punct initial
- Proces iterativ
  - Se executa salturi
  - Pana cand se gaseste minimul cu suficienta precizie
- O alta metoda presupune un alt mod de a calcula saltul



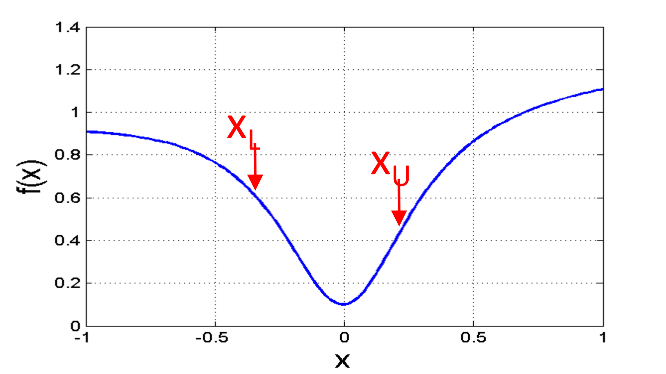
Metode de cautare

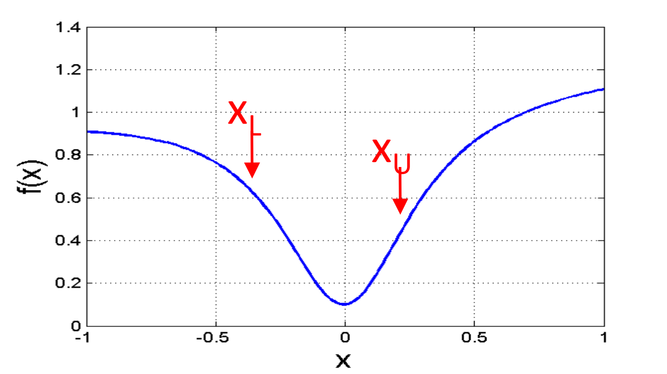
- Se porneste cu un interval inchis $[x_L, x_U]$ astfel incat sa includa minimul x^* .
- Se evaluaza $f(x)$ in doua puncte in interval .
- Se reduce intervalul.
- Se repeta procesul pana cand interval este destul de mic.



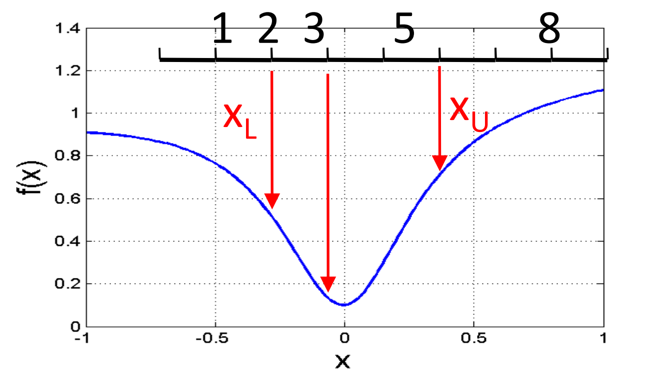
- Poate fi aplicata oricarei functii
  - Functia nu trebuie sa fie diferentiabila.

Metode de cautare





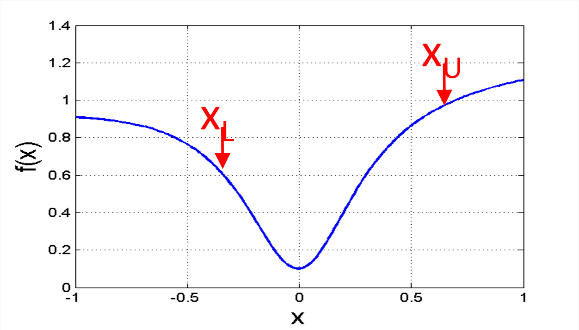
Dichotomous

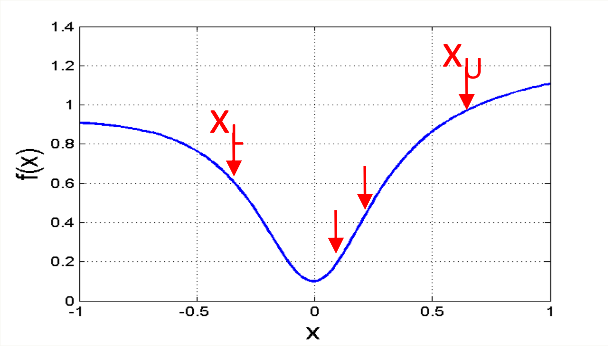


Fibonacci: 1 1 2 3 5 8 …

Metode de cautare

Dichotomous





Algoritm

0) Fie intervalul interval $[x_L, x_U] = [a,b]$

1) Calculam $x_1 = a + (b-a)/2 - E/2$ and $x_2 = a + (b-a)/2 + E/2$

E este rezolutia

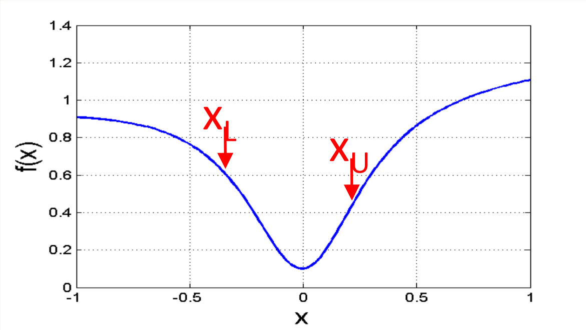
2) Se compara $f(x_1)$ cu $f(x_2)$

3) Daca $f(x_1) < f(x_2)$ atunci se elimina $x > x_2$ si se alege $b = x_2$

Daca $f(x_1) > f(x_2)$ atunci se elimina $x < x_1$ si se alege $a = x_1$

Daca $f(x_1) = f(x_2)$ atunci se alege alta pereche de puncte

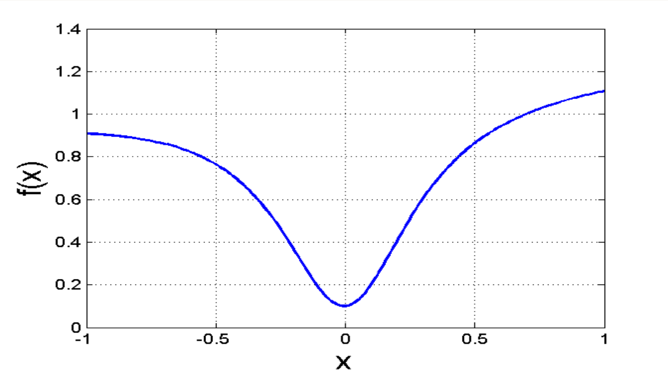
4) Se continua pana cand latimea intervalului este mai mica decat toleranta (2 E)

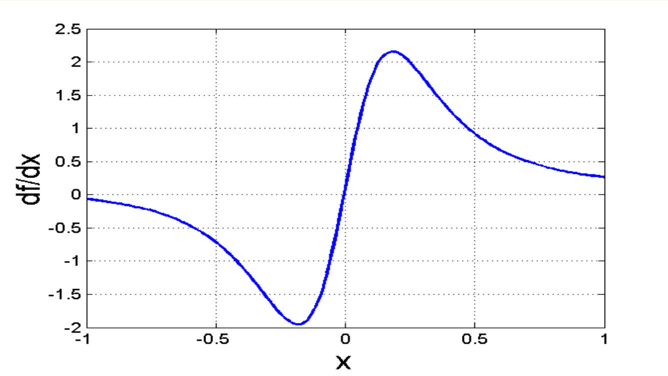


[Zitova]

Drept exemplu sa consideram functia

$$f(x) = 0.1 + 0.1x + x^2/(0.1 + x^2)$$



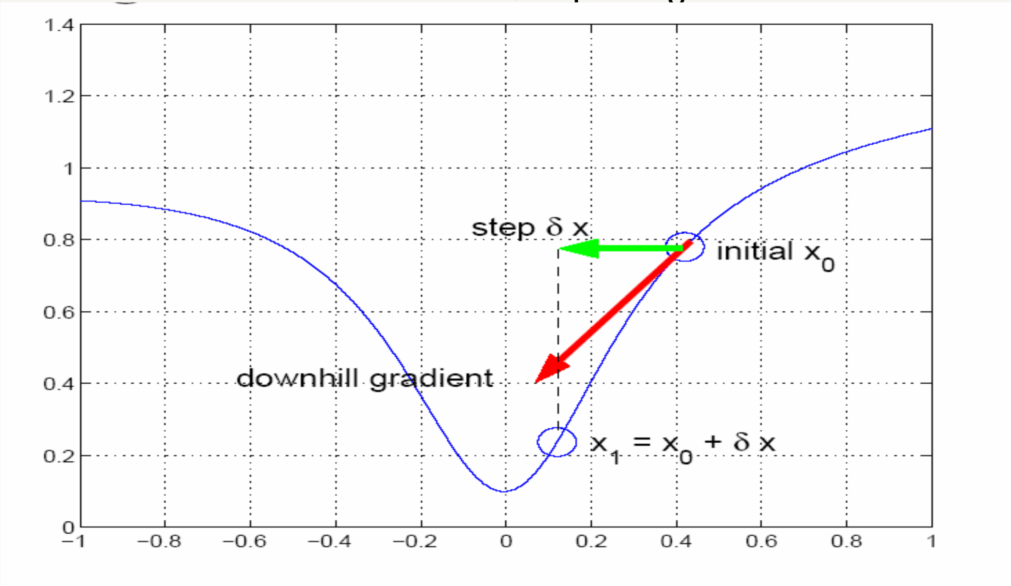


Metoda “Gradient descent”

Fiind data o locatie de start , x_0, se calculeaza gradientul df/dx

Si se merge in directia scaderii lui (la vale)

Acolo se genereaza un nou estimat, $x_1 = x_0 + \delta x$



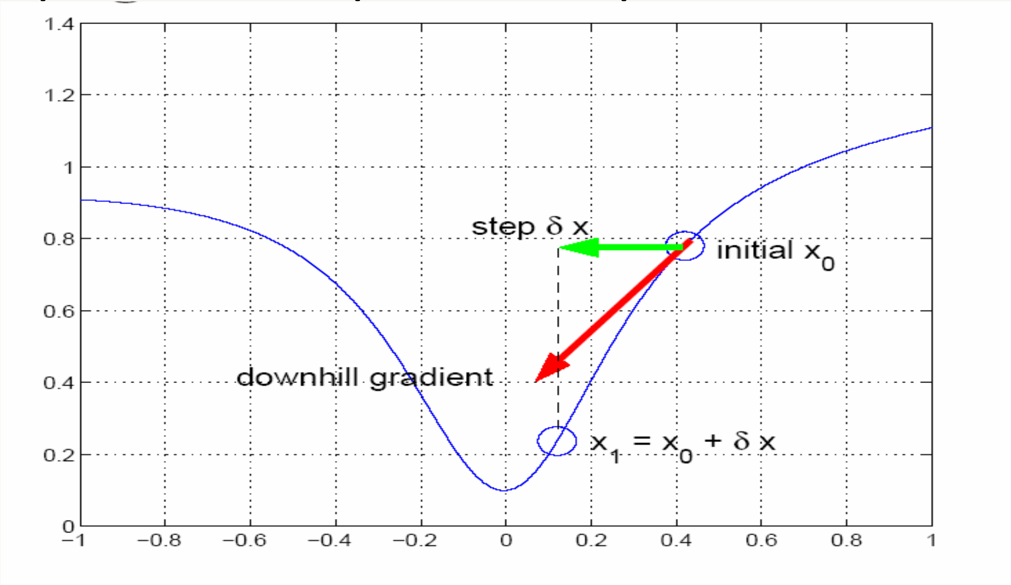
Cum se determina valoarea saltului δx?

Metoda “Gradient descent”

Fiind data o locatie de start , x_0, se calculeaza gradientul df/dx

Si se merge in directia scaderii lui (la vale)

Se alege saltul pe baza unui parametru prestabilit – learning rate **λ**



$\delta x = - \lambda * f'(x_0)$?

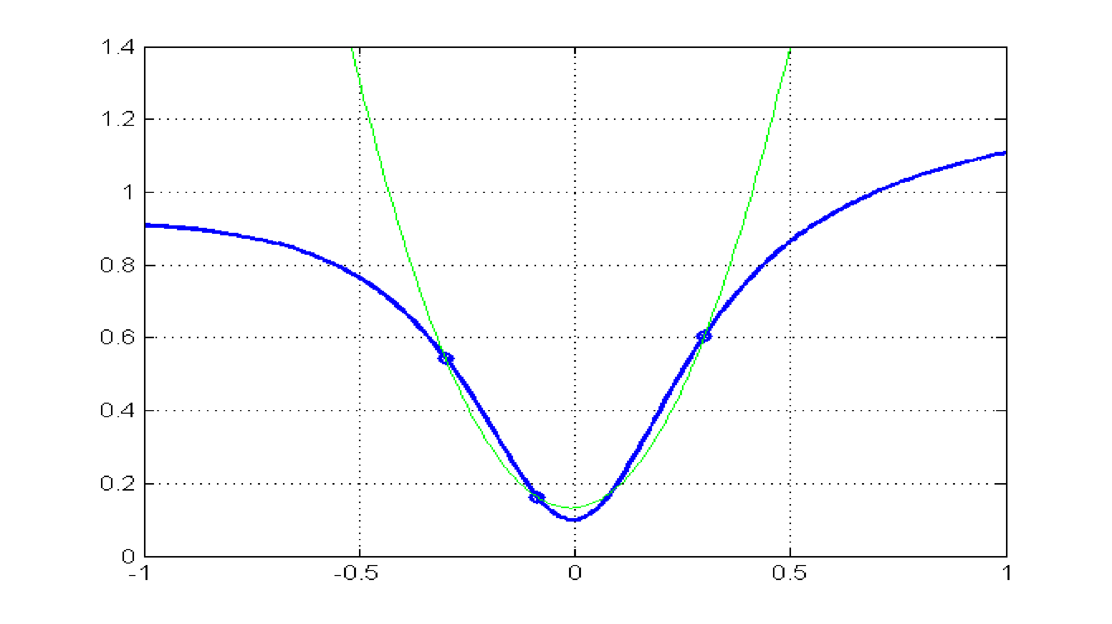
λ = 0.01

$x_1 = x_0 - \lambda * f'(x_0)$?

Interepretare polinomiala

- Se considera un interval care contine minimul.
- Se considera o aproximare cu un polinom de gradul 2 (aproximare patratica) sau de gradul 3 (cubica) care aproximeaza functia pe intervalul data.
- Pentru polinomul ales se calculeaza analitic pozitia minimului
- Noua valoare este data de minimul polinomul
- Repeta procesul.

Interpolare polinomiala



- Interpolare patratica folosind 3 puncte, 2 iteratii
-

Metode de tip Newton

In punctul curent se considera aproximarea functiei pe baze seriei Taylor cu 2 termeni.

- $$f(x+\delta x) = f(x) + f'(x)\delta x + \frac{1}{2} f''(x)\delta x^2 + o(\delta x^2)$$

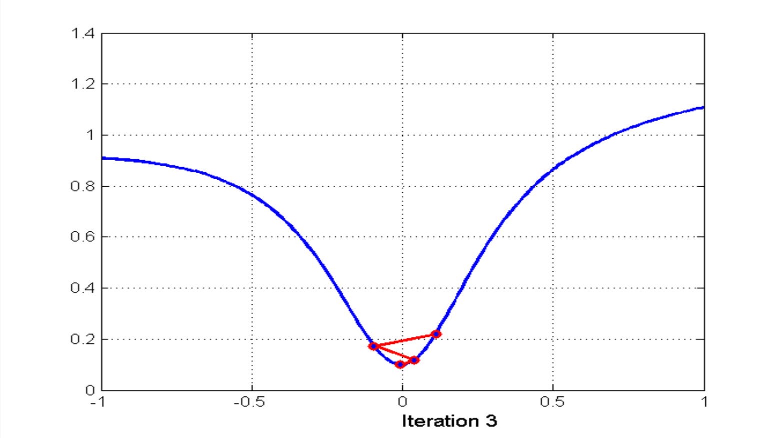
Se gaseste δx care minimizeaza aproximarea patratica

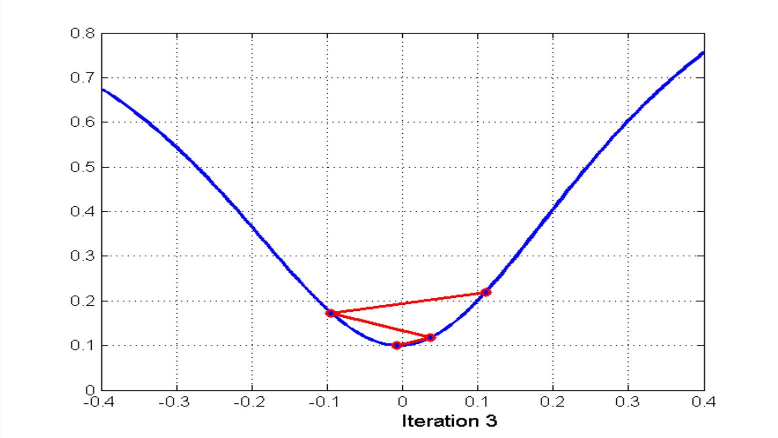
(polinom de gradul 2 – minim in -(b/2a)).

$$\delta x = -\frac{f'(x)}{f''(x)} \qquad x_{n+1} = x_n - \delta x = x_n - \frac{f'(x)}{f''(x)}$$

- Se considera noua valoare a lui x si se repeta procesul.

Metode de tip Newton

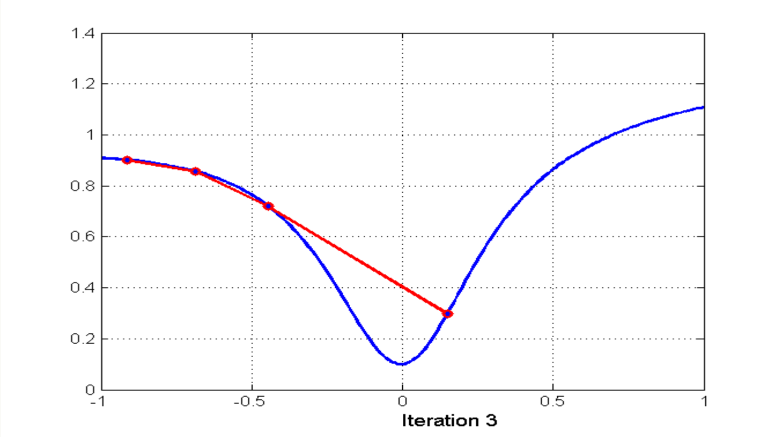


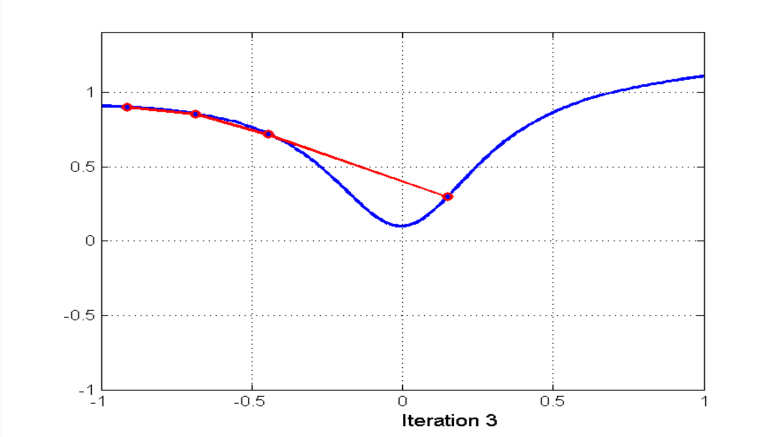


- Convergenta foarte rapida
- Necesita derivata a doua

Metode de tip Newton

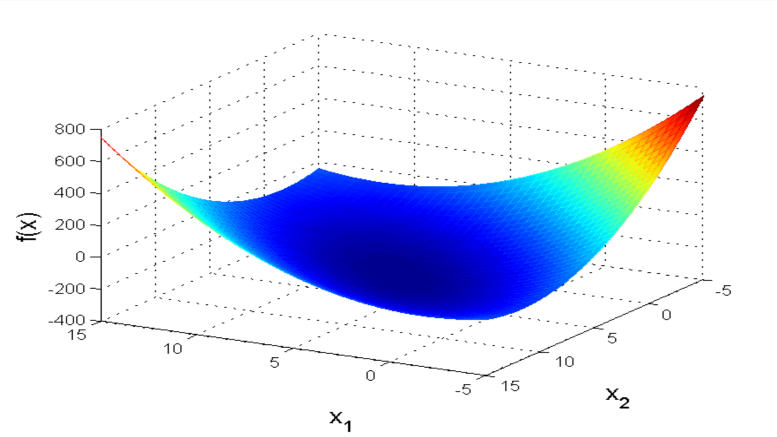
- Convergenta proasta pentru metode de tip Newton
  - Derivata a doua creeaza probleme
  - Probleme numerice.

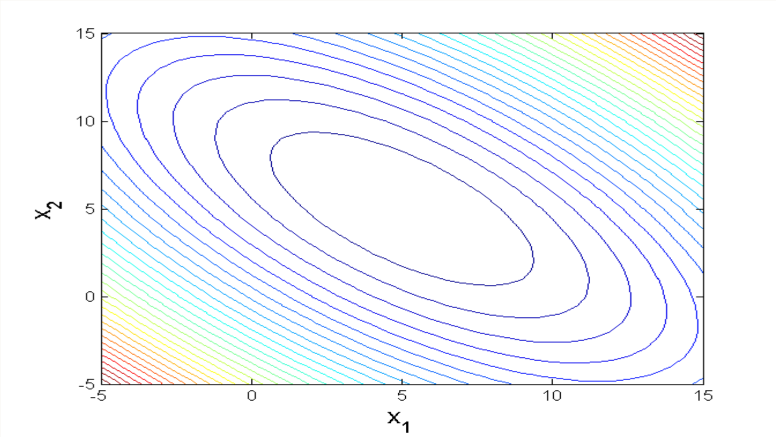




Extensia la spatii N-dimensionale (multivariate)

- N poate fi foarte mare
  - Retelele adanci au milioane de parametri
- In general se ilustreaza pe N=2 pentru vizualizare.





Algoritm generic de optimizare

- Start cu $\mathbf{x}_0$, $k = 0$.

1. Calculeaza o directie de cautare $\mathbf{p}_k$
2. Calculeaza o lungime a saltului α_k, astfel incat $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) < f(\mathbf{x}_k)$
3. Calculeaza noul pas $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$
4. Verifica convergenta (criteriu de stop) e.g. $df/d\mathbf{x} = \mathbf{0}$

$k = k+1$

Reduce optimizarea in N dimensiuni la o serie de minimizari liniare (1D)

Dezvoltare in serie Taylor

Dezvoltare multidimensionala in jurul unui punct $\mathbf{x}^*$

$$\boxed{f(\mathbf{x}^* + \mathbf{x}) \approx f(\mathbf{x}^*) + \nabla f^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \mathbf{H} \mathbf{x}}$$

Unde gradientul $\nabla f(\mathbf{x}^*)$ este un vector

$$\nabla f(\mathbf{x}^*) = \left[\frac{\partial f}{x_1} \dots \frac{\partial f}{x_N}\right]^T$$

Iar $\mathbf{H}(\mathbf{x}^*)$ este matricea hessian, simmetrica ce trebuie inversata

$$\mathbf{H}(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

Conjugate gradient

- Fiecare $\mathbf{p}_k$ este ales astfel incat sa fie perpendicular (conjugat) pe toate directiile de cautare precedente in raport cu Hessian $\mathbf{H}$:

$$\mathbf{p}_i^T \mathbf{H} \mathbf{p}_j = 0\,, \qquad i \neq j$$

- Directiile de cautare sunt mutual linear independente.
- *Remarcabil*, $\mathbf{p}_k$ poate fi ales doar cu cunostinte despre $\mathbf{p}_{k-1}$, , and

$$\nabla f(\mathbf{x}_{k-1}) \qquad \nabla f(\mathbf{x}_k)$$

$$\mathbf{p}_k = \nabla f_k + \left(\frac{\nabla f_k^\top \nabla f_k}{\nabla f_{k-1}^\top \nabla f_{k-1}} \right) \mathbf{p}_{k-1}$$

Metoda Newton

Se devolta $f(\mathbf{x})$ in serie Taylor in jurul punctului $\mathbf{x}_k$

$$f(\mathbf{x}_k + \delta\mathbf{x}) \approx f(\mathbf{x}_k) + \mathbf{g}_k^T \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T \mathbf{H}_k \delta\mathbf{x}$$

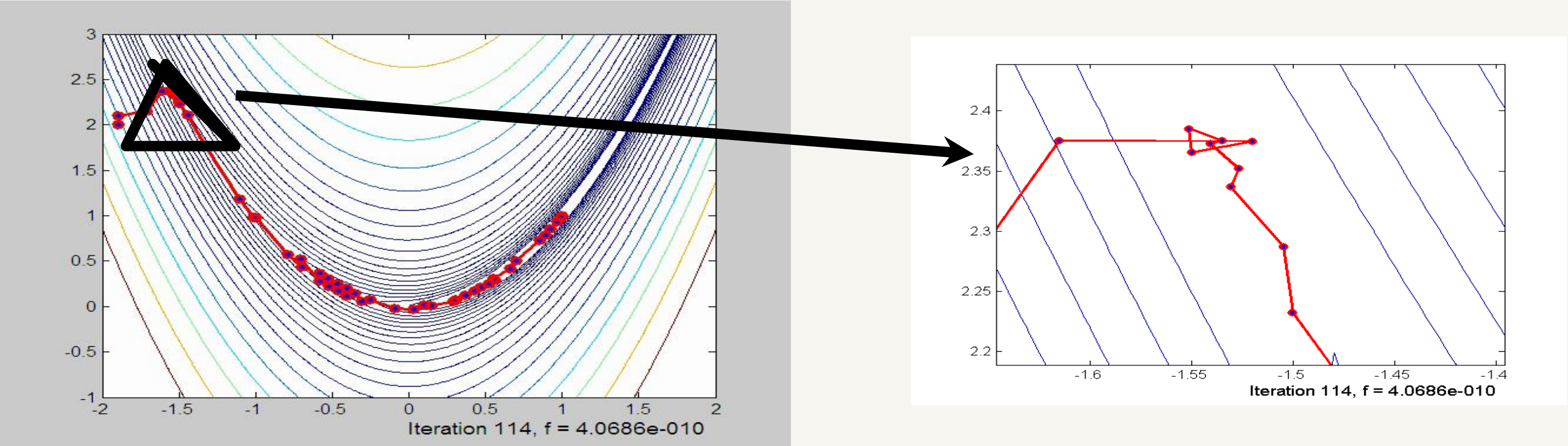
Unde gradientul este vectorul

$$\mathbf{g}_k = \nabla f(\mathbf{x}_k) = \left[\frac{\partial f}{x_1} \ldots \frac{\partial f}{x_N}\right]^T$$

Iar Hessiana este matricea simetrica

$$\mathbf{H}_k = \mathbf{H}(\mathbf{x}_k) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

Simplex



Optimizare cu Constrangeri

$$f(\mathbf{x}) : \mathbb{R}^N \longrightarrow \mathbb{R}$$

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x})$$

In functie de

- Constrangeri de tip egalitate : $a_i(\mathbf{x}) = 0 \quad i = 1, 2, \ldots, p$
- Constrangeri de tip inegalitate: $c_j(\mathbf{x}) \geq 0 \quad j = 1, 2, \ldots, q$

- Constrangerile definesc o reziune fezabile nevida.
- Idea este sa o convertim intr-o optimizare fara constrangeri.

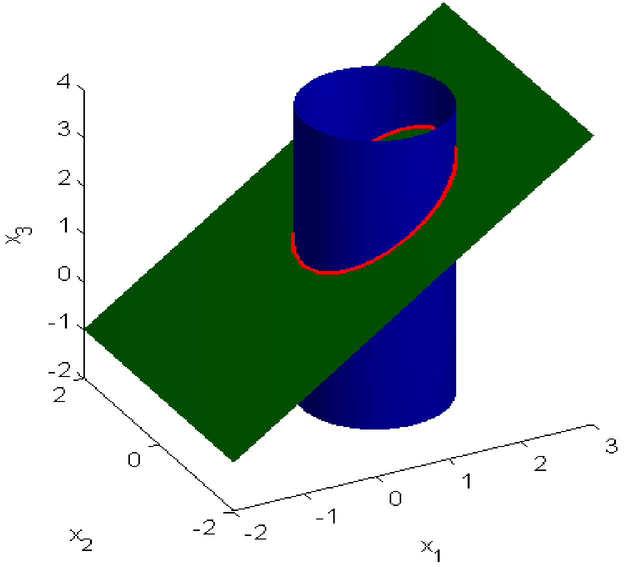
Constrangeri de tip egalitate

- Minimizam $f(\mathbf{x})$ cu constrangerea : $a_i(\mathbf{x}) = 0$ $i = 1, 2, \ldots, p$
  pentru

- Gradientul lui $f(\mathbf{x})$ intro zona restransa este egala cu combinatia liniara a gradientilor lui $a_i(\mathbf{x})$ de inmultit cu multiplicatorii lui Lagrange drept coeficienti.

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*)$$

Exemplu 3D

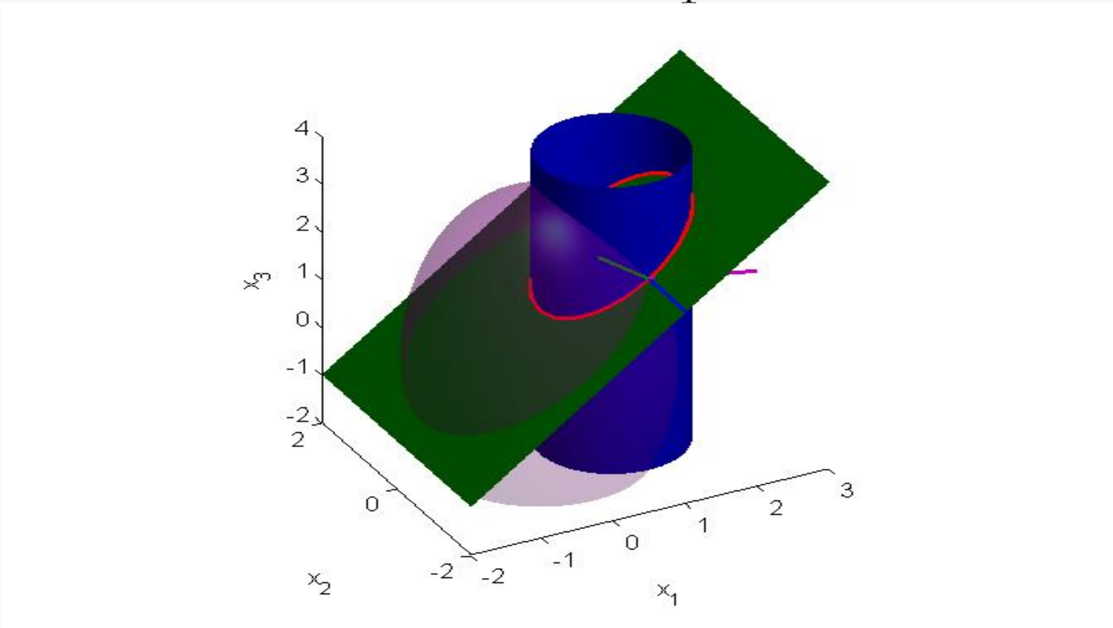


$$\boxed{a_1(\mathbf{x}) = -x_1 + x_3 - 1 = 0}$$

$$\boxed{a_2(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1 = 0}$$

Exemplu 3D

$$f(\mathbf{x}) = x_1^2 + x_2^2 + \frac{1}{4}x_3^2$$

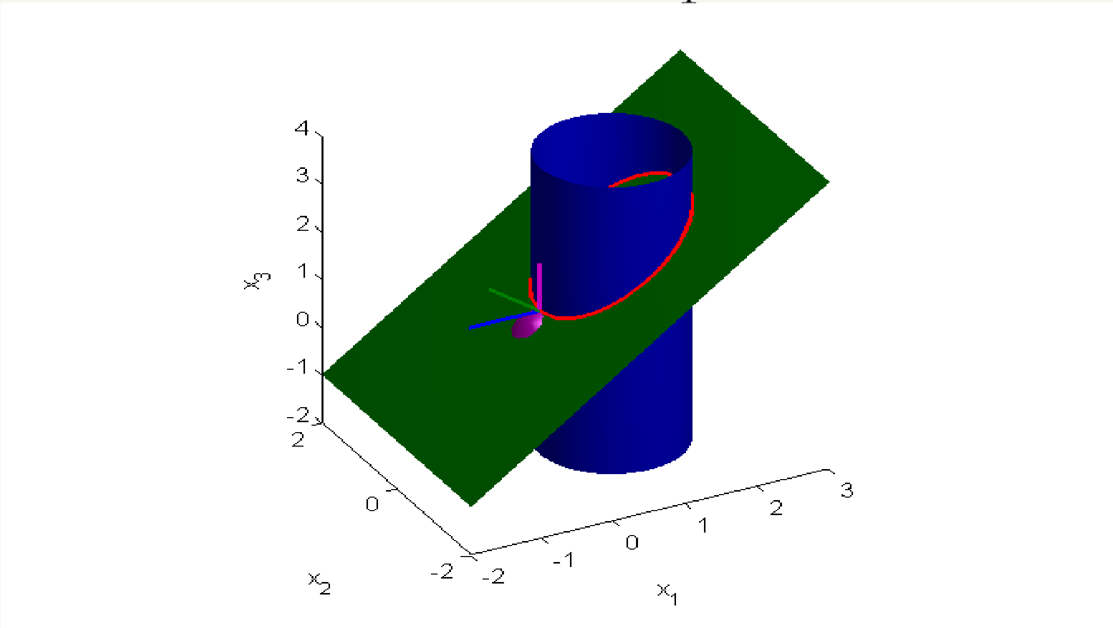


$$f(\mathbf{x}) = 3$$

Gradients of constraints and objective function are linearly independent.

Exemplu 3D

$$f(\mathbf{x}) = x_1^2 + x_2^2 + \frac{1}{4}x_3^2$$



$$f(\mathbf{x}) = 1$$

Gradients constrangerii si functiei obiective sunt lineari dependenti.