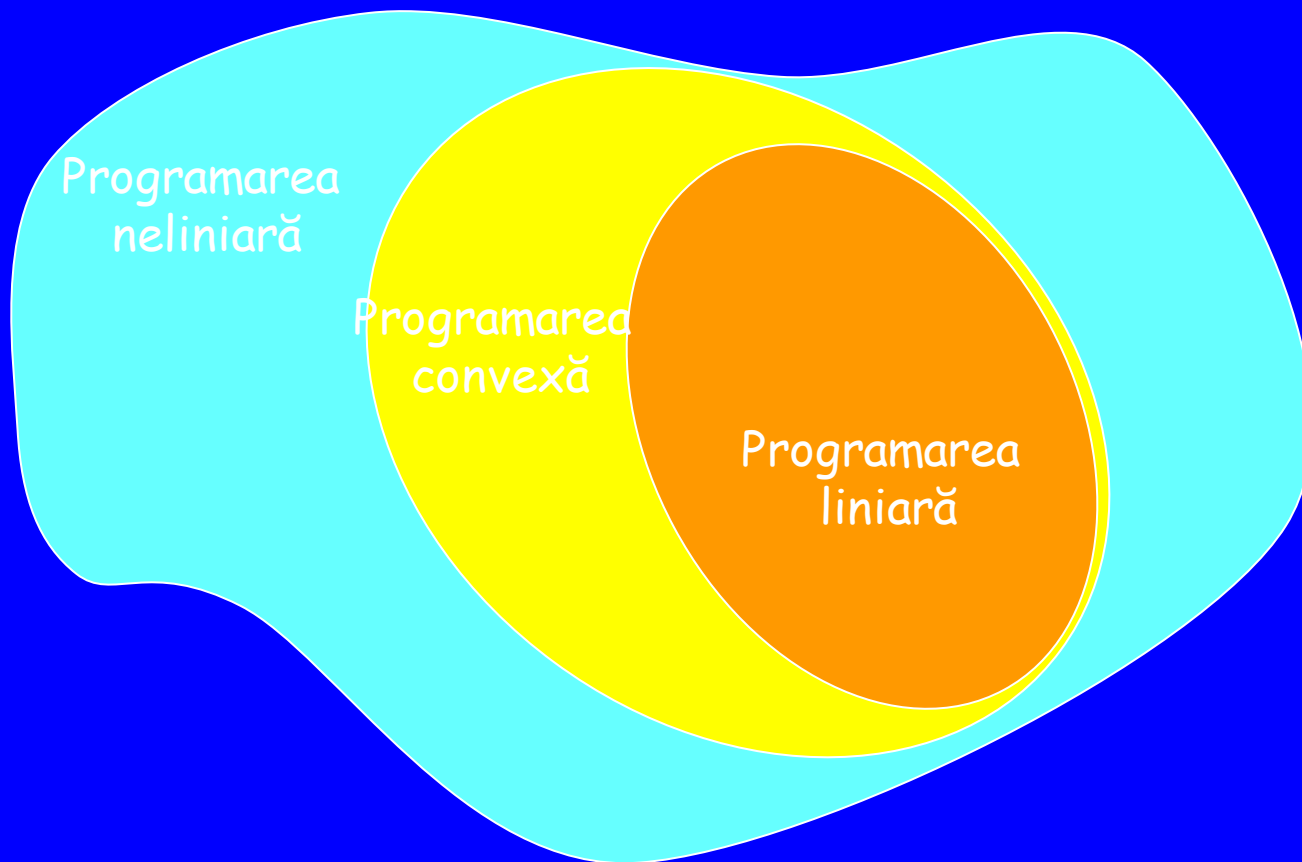


Programarea neliniară

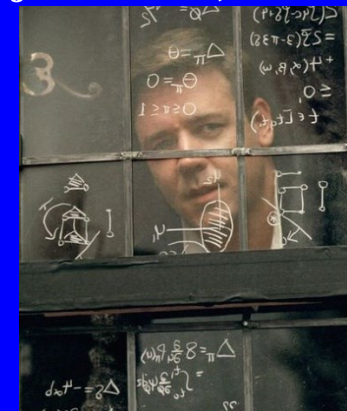
Ierarhizarea problemelor de optimizare



1. Introducere

Instoric:

- **Harold William Kuhn (1925 -)**
(Condițiile necesare și suficiente pentru soluția optimală a problemelor de programare neliniară - 1951, teoria jocurilor)
- **Albert William Tucker (1905-1995)**
(Condițiile necesare și suficiente pentru soluția optimală a problemelor de programare neliniară – 1951, teoria jocurilor)



Modelul matematic general al unei probleme de programare liniară

$$\min (\max) F(\mathbf{X})$$

$$g_i(\mathbf{X}) (\leq, \geq, =) b_i, \quad i = 1, \dots, m$$

$\mathbf{X} = (x_1, \dots, x_n)^t$ vectorul n-dimensional al variabilelor de optimizare

$F(\mathbf{X})$ – funcția obiectiv;

$g_i(\mathbf{X})$ restricțiile modelului matematic;

b_i – vectorul termenilor liberi.

Programarea neliniară fără restricții

- Pentru funcțiile obiectiv cu o singură variabilă pentru determinarea soluției optime se folosesc derivatele de ordinul 1 și 2.
- Pentru funcțiile cu mai multe variabile pentru determinarea soluției optime se folosesc gradientul și matricea Hessiană.
- Gradientul unei funcții reprezintă derivata de ordinul 1 în raport cu toate variabilele, iar matricea Hessiană este reprezentată de derivatele de ordinul 2.

Gradientul unei funcții

Pentru o funcție F definită în raport cu variabilele x_1, x_2, \dots, x_n , gradientul este definit astfel:

$$\nabla F = \left[\frac{\partial F}{\partial x_1} \quad \frac{\partial F}{\partial x_2} \quad \dots \quad \frac{\partial F}{\partial x_n} \right]$$

Exemplu: $F = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$

$$\nabla F = \left[15 - 3(x_3)^2 \quad 6(x_2)^2 \quad -6x_1x_3 \right]$$

Matricea Hessiană

- Matricea Hessiană (∇^2) a unei funcții $F(x_1, x_2, \dots, x_n)$ este definită astfel:

$$\nabla^2 F = \begin{bmatrix} \frac{\partial F^2}{\partial x_1^2} & \frac{\partial F^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial F^2}{\partial x_1 \partial x_n} \\ \frac{\partial F^2}{\partial x_2 \partial x_1} & \frac{\partial F^2}{\partial x_2^2} & \cdots & \frac{\partial F^2}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F^2}{\partial x_n \partial x_1} & \frac{\partial F^2}{\partial x_n \partial x_2} & \cdots & \frac{\partial F^2}{\partial x_n^2} \end{bmatrix}$$

Matricea Hessiană

- Exemplu:

$$F = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$$

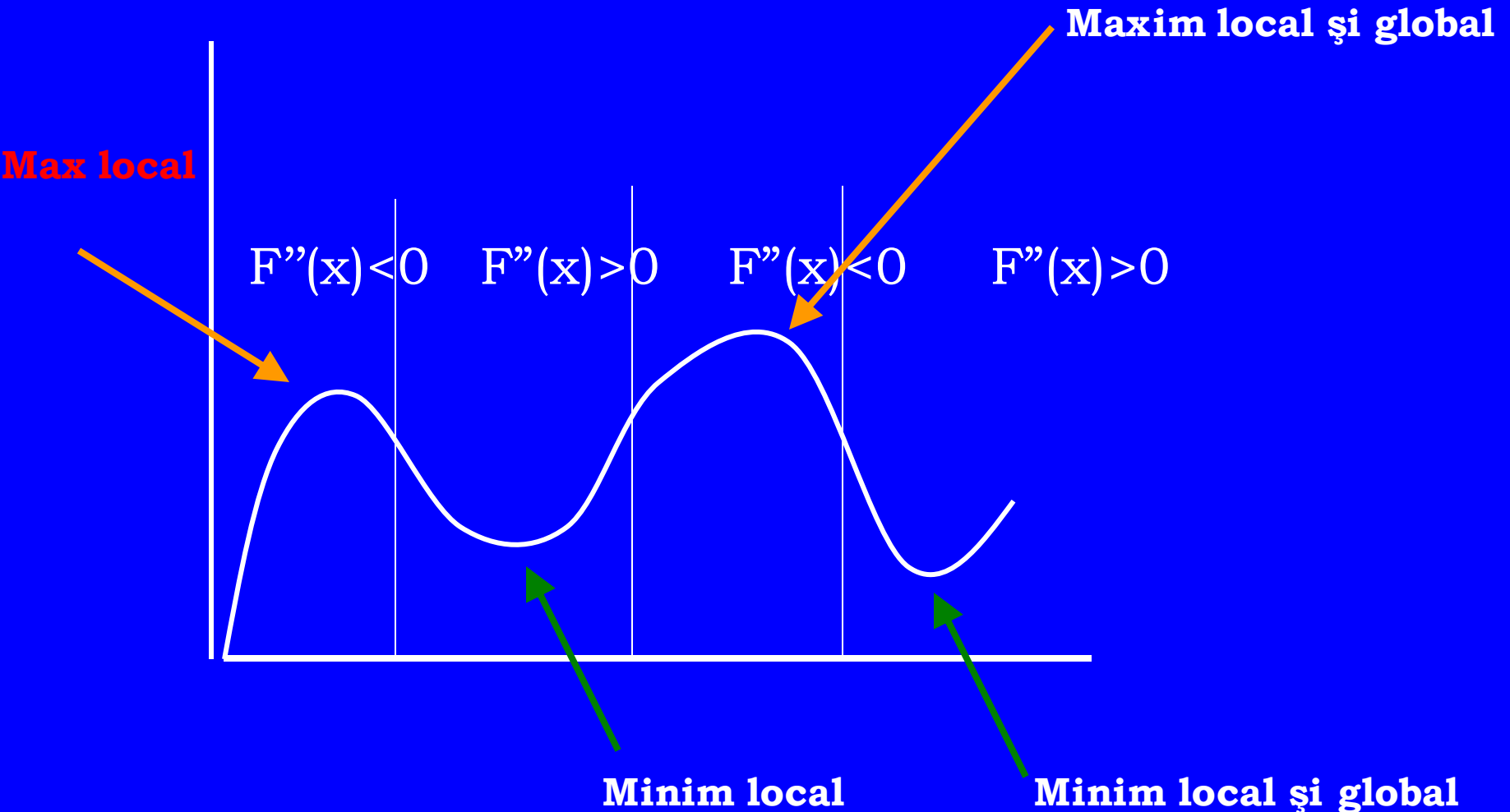
$$\nabla_j F = [15 - 3(x_3)^2 \quad 6(x_2)^2 \quad -6x_1x_3]$$

$$\nabla^2 F = \begin{bmatrix} 0 & 0 & -6x_3 \\ 0 & 12x_2 & 0 \\ -6x_3 & 0 & -6x_1 \end{bmatrix}$$

Condițiile de optimalitate

- Optimizarea fără restricții – **problemă de calcul cu mai multe variabile**. Pentru $Y=F(X)$, punctul de optim va fi punctul în care $F'(X) = 0$ iar $F''(X)$ îndeplinește condițiile privind derivata de ordinul al II-lea.
- Punctul de minim îndeplinește condițiile:
 $F'(X) = 0$ și $F''(X) > 0$.
- Punctul de maxim îndeplinește condițiile:
 $F'(X) = 0$ and $F''(X) < 0$.

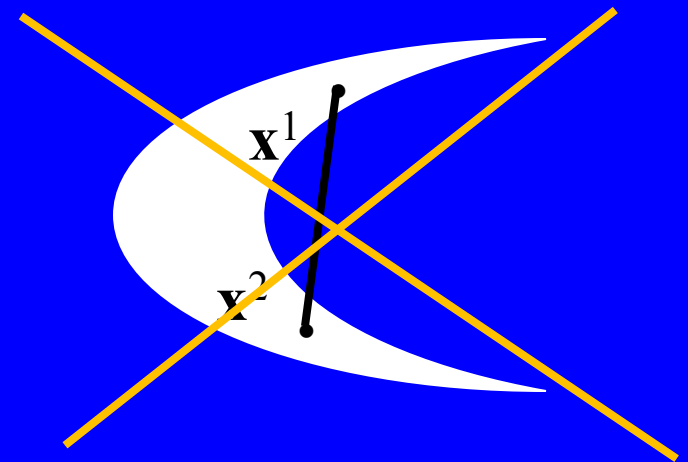
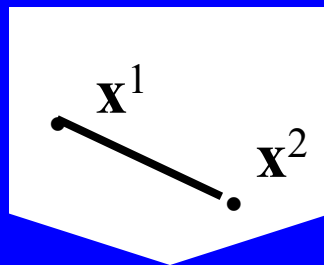
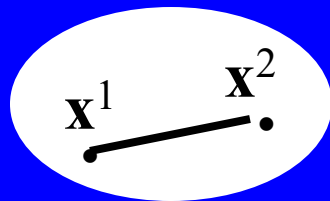
Concavitătea și derivata de ordinul II



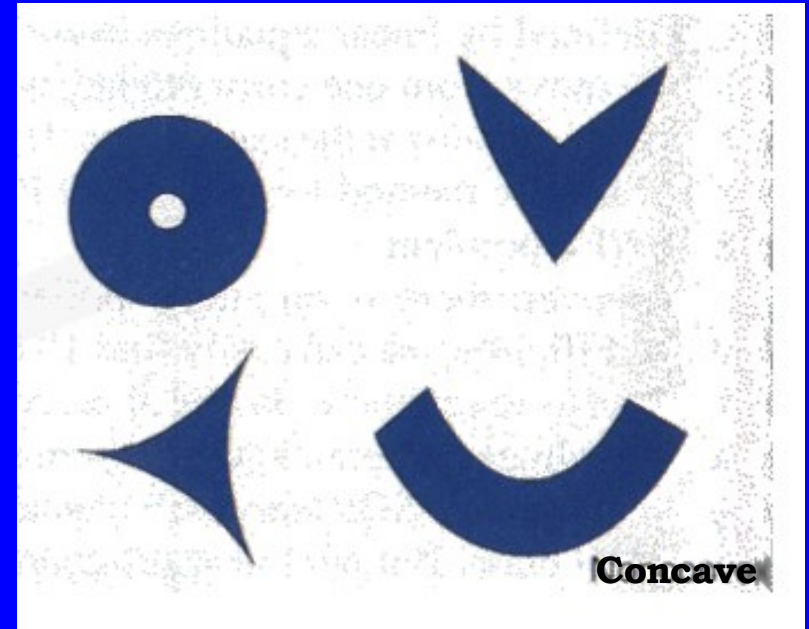
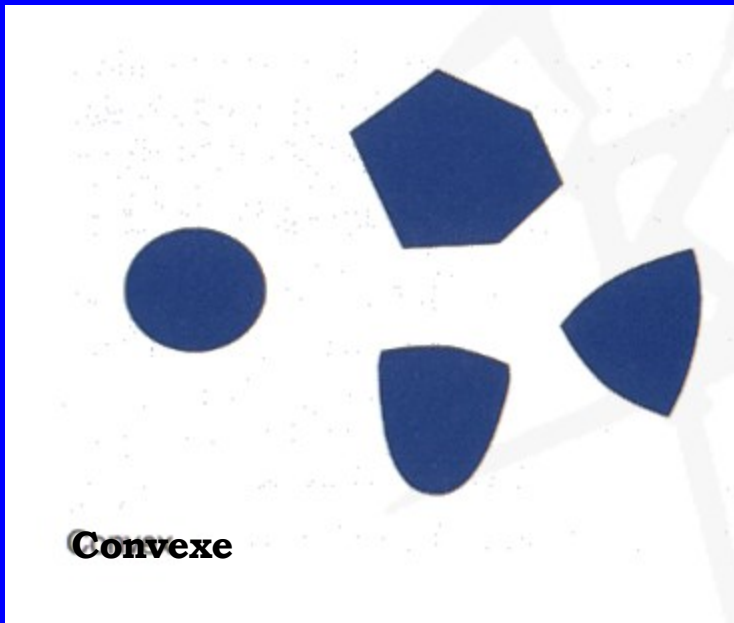
Mulțimi convexe

Definiție: O mulțime $S \subseteq \mathbb{R}^n$ este convexă dacă orice punct de pe segmentul de dreaptă care unește oricare două puncte $x^1, x^2 \in S$ aparține mulțimii S . Din punct de vedere matematic putem scrie:

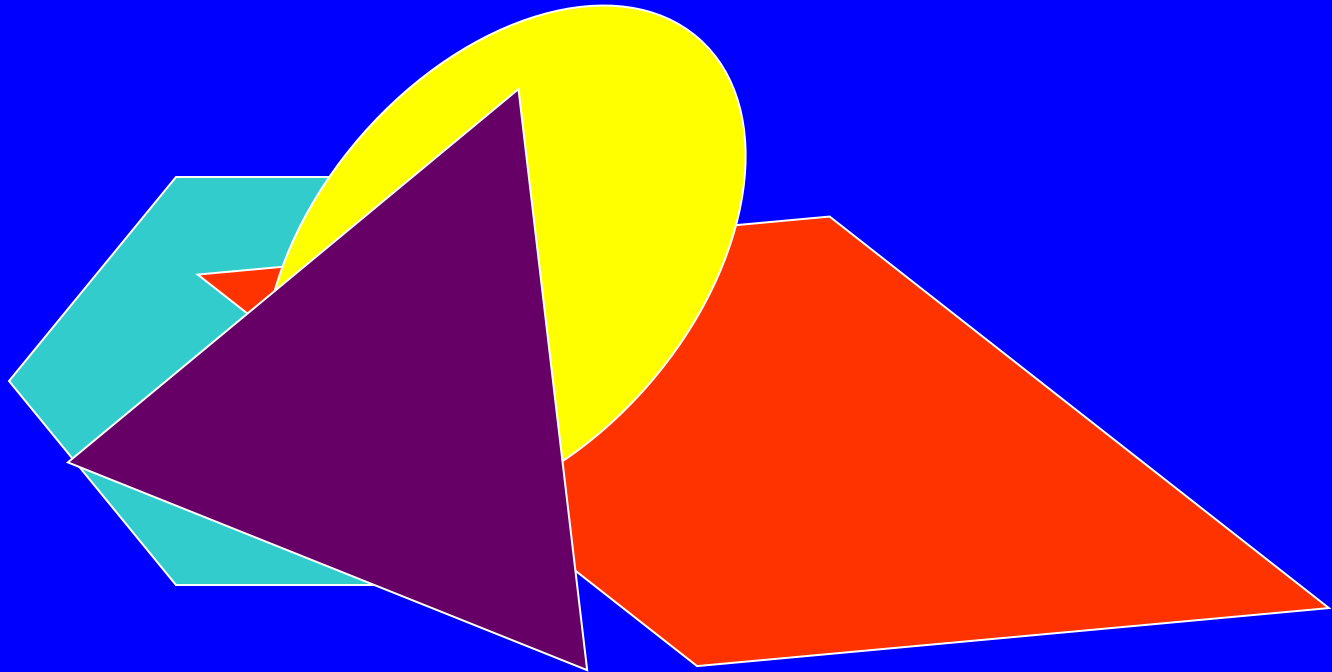
$$x^0 = \lambda x^1 + (1-\lambda)x^2 \in S \text{ pentru } \lambda \in [0, 1].$$



Mulțimi convexe și concave



Mulțimi convexe



Intersecția unui număr oricât de mare de mulțimi convexe este o mulțime convexă.

Exemplu:

$$\max (x_1 - x_2)$$

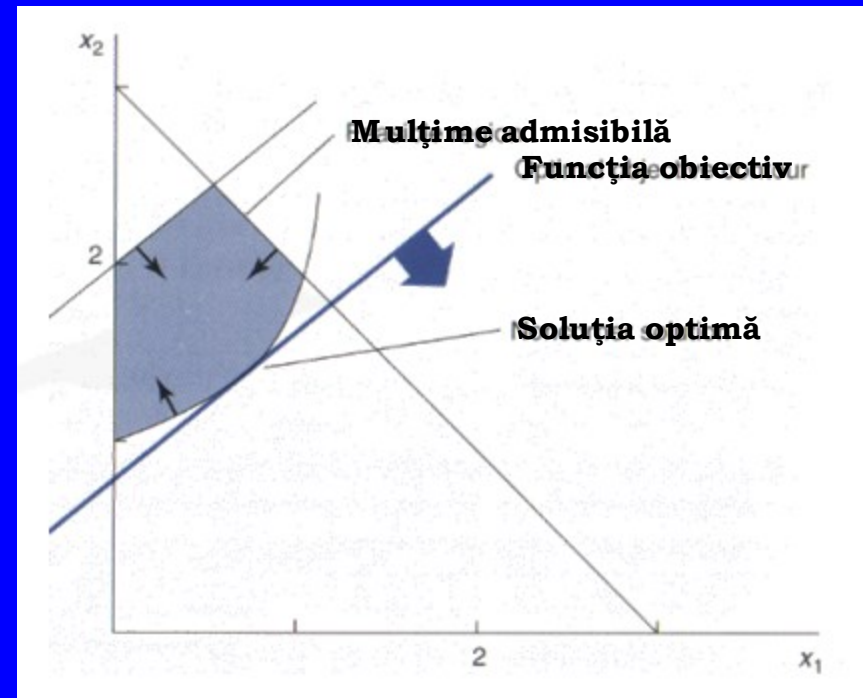
$$- x_1^2 - x_2 \geq 1$$

$$x_1 + x_2 \leq 3$$

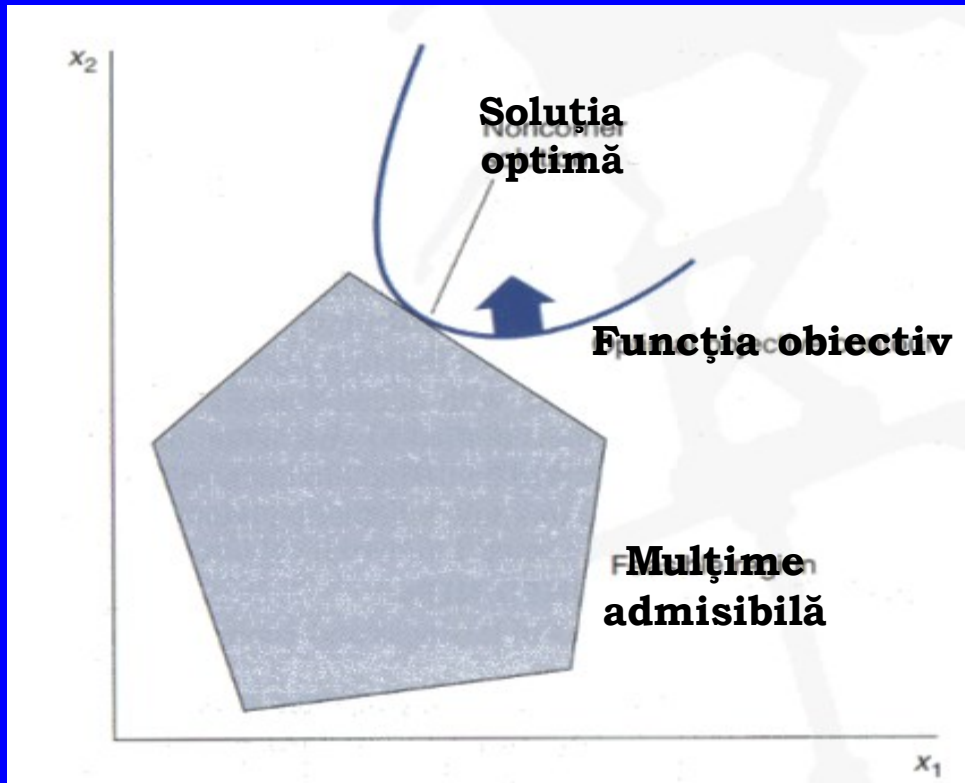
$$- x_1 + x_2 \geq 2$$

$$x_1 \geq 0, x_2 \geq 0$$

- **Muțimea admisibilă a soluțiilor nu este un poliedru convex.**
- **Soluția optimă nu se găsește în vârful unui poliedru convex**



Funcție obiectiv neliniară cu restricții liniare.

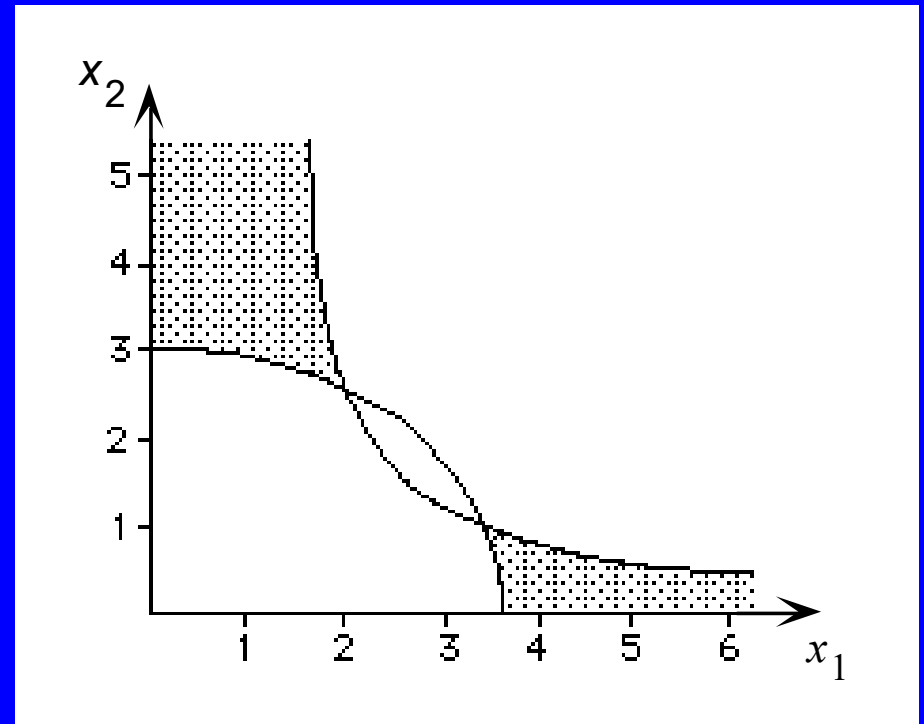


Deși mulțimea admisibilă este un poliedru convex, soluția optimă nu se găsește într-unul din vârfurile acestui poliedru.

Există și situații în care soluția optimă se găsește în interiorul Mulțimii admisibile.

Exemplu de mulțime admisibilă neconvexă

$$S = \{(x_1, x_2) : (0.5x_1 - 0.6)x_2 \leq 1$$
$$2(x_1)^2 + 3(x_2)^2 \geq 27; x_1, x_2 \geq 0\}$$

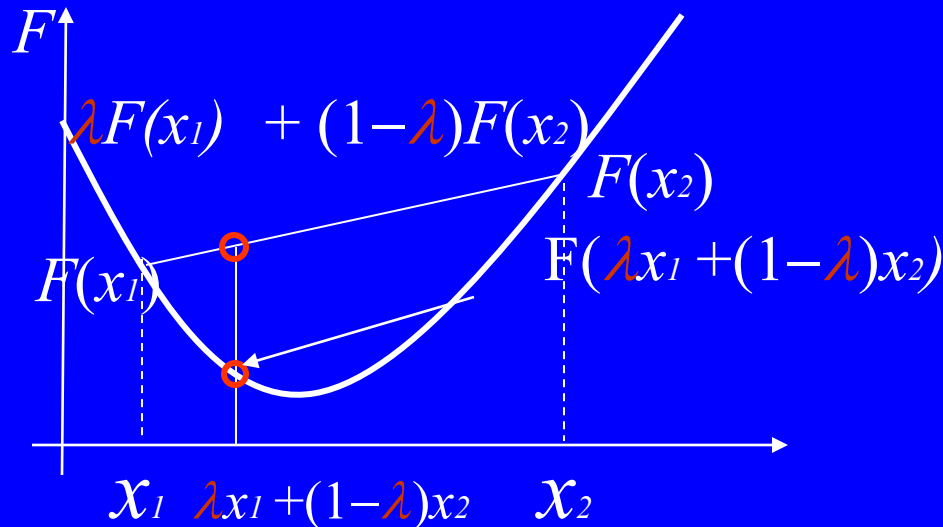


Funcție convexă

$S \subseteq \mathbb{R}^n$ — Mulțime convexă

$F : S \rightarrow \mathbb{R}$ — este o funcție convexă dacă

Dacă este valabilă inegalitatea atunci funcția $F(X)$ este strict convexă.

$$F(\lambda x_1 + (1-\lambda)x_2) \leq \lambda F(x_1) + (1-\lambda)F(x_2), 0 \leq \lambda \leq 1$$


Programarea convexă

Fie $S = \{ X \in \mathfrak{R}^n : g_i(X) \leq b_i, i = 1, \dots, m \}$

Definiție: Dacă $g_i(X)$ este o funcție convexă pentru orice $i = 1, \dots, m$ atunci S este o mulțime convexă.

Teorema programării convexe: Fie $x \in \mathfrak{R}^n$ și $F(X)$ o funcție convexă definită pe o mulțime convexă de restricții S . Dacă există o soluție finită pentru problema:

$$\min \{ F(X) : X \in S \}$$

Atunci toate punctele de optim local sunt și puncte de optim global. Dacă $F(X)$ este strict convexă, punctul de optim este unic.

Programarea convexă

$$\text{Min } F(x_1, \dots, x_n)$$

$$\begin{aligned} g_i(x_1, \dots, x_n) &\leq b_i \\ i &= 1, \dots, m \\ x_1 &\geq 0, \dots, x_n \geq 0 \end{aligned}$$

este o problemă de programare convexă dacă F este convexă și fiecare restricție g_i este convexă.

$$\text{Max } F(x_1, \dots, x_n)$$

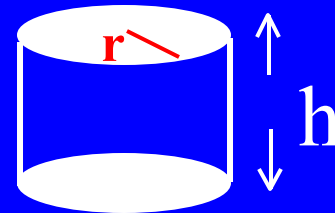
$$\begin{aligned} g_i(x_1, \dots, x_n) &\leq b_i \\ i &= 1, \dots, m \\ x_1 &\geq 0, \dots, x_n \geq 0 \end{aligned}$$

Este o problemă de programare convexă dacă F este concavă și fiecare restricție g_i este convexă.

Exemplu:

Să se construiască un cilindru (având și părțile superioară și inferioară) având un volum maxim, astfel încât suprafața sa să nu depășească o coală de tablă de dimensiune s .

$$\begin{aligned} \max \quad & V(r, h) = \pi r^2 h \\ & 2\pi r^2 + 2\pi r h = s \\ & r \geq 0, h \geq 0 \end{aligned}$$



Există mai multe moduri de abordare a acestei probleme. Unul dintre acestea determină înălțimea h din restricția de suprafață și înlocuiește expresia obținută în funcția obiectiv.

Soluția

$$h = \frac{s - 2\pi r^2}{2\pi r} \rightarrow \text{Volumul} = V = \pi r^2 \left[\frac{s - 2\pi r^2}{2\pi r} \right] = \frac{rs}{2} - \pi r^3$$

$$\frac{dV}{dr} = 0 \Rightarrow r = \left(\frac{s}{6\pi}\right)^{1/2} \quad h = \frac{s}{2\pi r} - r = 2\left(\frac{s}{6\pi}\right)^{1/2}$$

$$V = \pi r^2 h = 2\pi \left(\frac{s}{6\pi}\right)^{3/2} \quad r = \left(\frac{s}{6\pi}\right)^{1/2} \quad h = 2\left(\frac{s}{6\pi}\right)^{1/2}$$

Este soluția obținută soluție optimă?

Testul de convexitate

$$V(r) = \frac{rs}{2} - \pi r^3 \rightarrow \frac{dV(r)}{dr} = \frac{s}{2} - 3\pi r^2 \rightarrow \frac{d^2V(r)}{dr^2} = -6\pi r$$

$$\frac{d^2V}{dr^2} \leq 0 \text{ pentru orice } r \geq 0$$

$V(r)$ este concavă pentru orice $r \geq 0$ astfel încât soluția obținută reprezintă maximul global.

Teoreme

Teorema 1.

Pentru o funcție $F(X)$, derivabilă pe o mulțime convexă, sunt valabile următoarele afirmații echivalente:

1. $F(X)$ este convexă;

2. $F(X_2) - F(X_1) \geq (\nabla F(X_1), X_2 - X_1)$, oricare ar fi $X_2, X_1 \in S$
adică valoarea funcției $F(X)$ crește pe direcția gradientului $\nabla F(X_1)$.
 (X, Y) este produsul scalar al vectorilor X și Y .

3. Produsul scalar $(\nabla F(X + \lambda d), d)$ este nedescrescător în raport cu λ .

4. Matricea Hessiană este semi-pozitiv definită pentru toate punctele $X \in S$, adică:

$$(d, H(X) \cdot d) \geq 0$$

În cazul funcțiilor strict convexe, pentru punctele 2 și 3 sunt valabile doar inegalitățile

Teorema 2.

Orice punct de minim local este și punct de minim global. Dacă funcția $F(X)$ este strict convexă atunci soluția optimală (dacă există) este unică.

Teorema 3.

Se consideră următoarea problemă de programare convexă:

$$\min (F(X)), \quad X \in S$$

unde: S este mulțimea admisibilă de soluții (este o mulțime convexă);
 $F(X)$ – derivabilă și convexă;

X^* reprezintă o soluție a problemei convexe de mai sus dacă și numai dacă sunt adevărate următoarele :

$$(\nabla F(X^*), X - X^*) \geq 0, \text{ oricare ar fi } X \in S$$

In cazul în care $X \in \mathbb{R}$, iar restricțiile lipsesc relația devine:

$$\nabla F(X^*) \geq 0$$

Dacă $F(X)$ este derivabilă și convexă iar mulțimea este strict convexă, atunci X^* este unic.

Teorema 3 prezintă condițiile necesare și suficiente pentru un punct de minim.

Programarea neliniară convexă fără restricții

Se consideră următoarea problemă de programare convexă:

$$\min (F(X)), X \in S \subseteq R^n$$

Condițiile necesare și suficiente ca un punct X^* să fie un minim local și global sunt:

$$\nabla F(X^*) = 0 \text{ și } (d, H(X)d) \geq 0$$

iar pentru un minim global unic, acestea sunt:

$$\nabla F(X^*) = 0 \text{ și } (d, H(X)d) > 0$$

Notății: $g^{(k)} = \nabla F(X^{(k)})$, $H^{(k)} = \nabla^2 F(X^{(k)})$

Programarea neliniară convexă fără restricții

În general, pentru rezolvarea unei probleme de programare neliniară se folosește relația iterativă:

$$X^{(k+1)} = X^{(k)} + \lambda^{(k)} \cdot d^{(k)}$$

unde:

$d^{(k)} \in R^n$ – reprezintă o direcție de deplasare în iterația k din punctul curent $X^{(k)}$;

$\lambda^{(k)}$ – scalar ce reprezintă lungimea pasului de deplasare în iterația curentă.

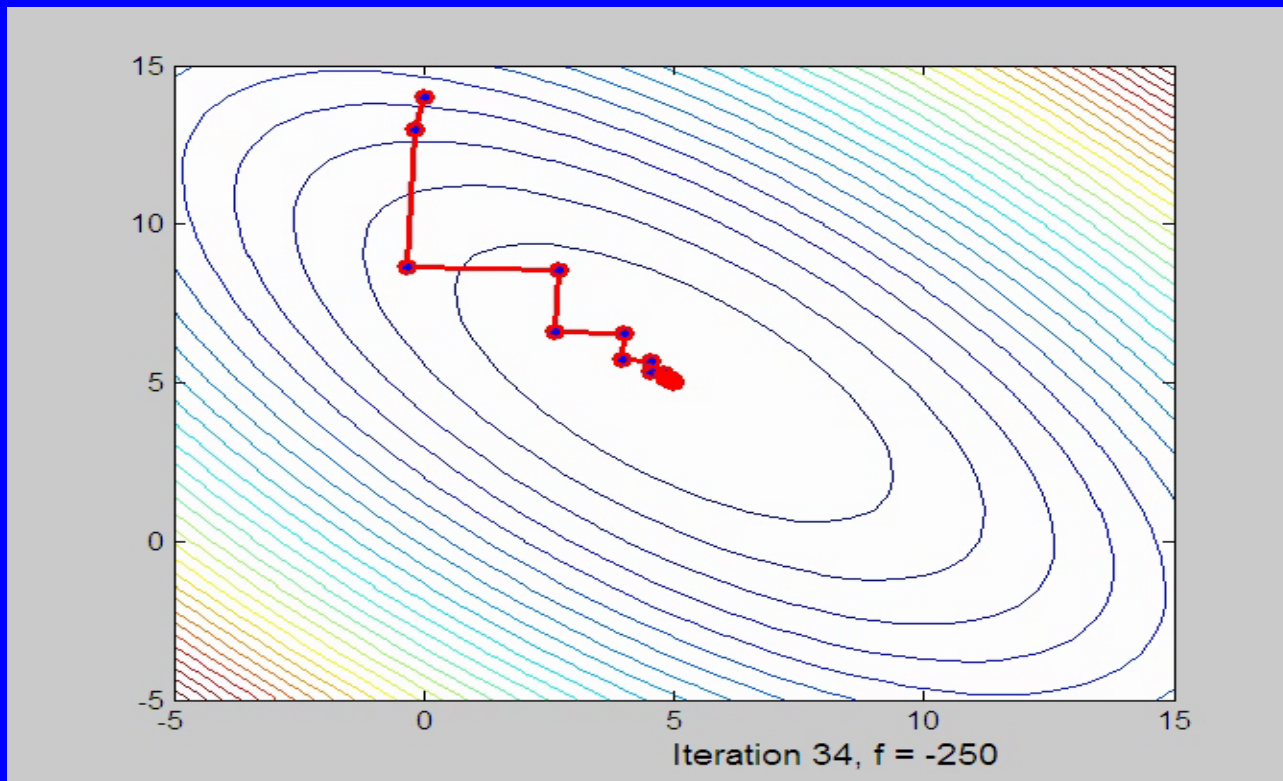
OBSERVAȚII:

- 1. Metodele de optimizare diferă prin procedurile concrete de alegere a parametrilor $d^{(k)}$ și $\lambda^{(k)}$.**
- 2. Având în vedere reducerea efortului de calcul privind implementarea procedurii iterative, în fiecare etapă, informația disponibilă pentru obținerea direcției $d^{(k)}$ și a pasului $\lambda^{(k)}$ este strict limitată la valorile funcției și a primei sale derivate.**

Proceduri pentru alegerea pasului de deplasare

Determinarea pasului de deplasare $\lambda^{(k)}$ constituie o parte critică a metodelor de optimizare din cadrul programării neliniare.

O **valoare prea mică** a pasului conduce la un **număr mare de iterații** în procesul de optimizare, deci la **creșterea timpului de calcul**, iar o **valoare prea mare** poate conduce la **oscilații în jurul minimumului** și deci, din nou, la **creșterea timpului de calcul**.



- **Gradientul este întotdeauna perpendicular pe normala la suprafața funcției obiectiv.**
-
- **După fiecare iterație noua direcție a gradientului este întotdeauna perpendiculară pe direcția din iterația anterioară.**
-
- **În consecință, iterațiile au tendința de a asigura o deplasare în zig-zag într-un mod foarte ineficient.**

Metode de alegere optimală a pasului de deplasare

Aceste metode urmăresc determinarea pasului $\lambda^{(k)}$
Din condiția realizării minimumului funcției $F(X)$
pe direcția $d^{(k)}$:

$$F\left(X^{(k)} + \lambda^{(k)} d^{(k)}\right) = \min F\left(X^{(k)} + \lambda d^{(k)}\right) = \min F(\lambda)$$

Dacă se face ipoteza că funcția $F(\lambda)$ este netedă,
metodele cele mai eficiente se bazează pe
aproximarea acesteia printr-o funcție polinomială
de gradul 2, al cărei minim se poate calcula ușor.

$$F = C_2 \lambda^2 + C_1 \lambda + C_0, \quad C_2 > 0$$

cu minimumul

$$\lambda^{(k)} = -\frac{C_1}{2C_2}$$

Metoda 1

Pentru determinarea coeficienților C_0 , C_1 și C_2 sunt necesare trei valori ale funcției respective în punctele λ_1 , λ_2 și λ_3 , expresia finală fiind:

$$\lambda^{(k)} = \frac{1}{2} \cdot \frac{F_1(\lambda_2^2 - \lambda_3^2) + F_2(\lambda_3^2 - \lambda_1^2) + F_3(\lambda_1^2 - \lambda_2^2)}{F_1(\lambda_2 - \lambda_3) + F_2(\lambda_3 - \lambda_1) + F_3(\lambda_1 - \lambda_2)}$$

- Corectitudinea minimumului depinde de aproximarea polinomială;
- Eficiența procesului de optimizare depinde foarte mult de alegerea valorilor λ_2 și λ_3 (pentru λ_1 se ia, de regulă, o valoare nulă).

Metoda 2

Constă în folosirea valorii funcției și a derivatei acesteia în punctul curent $X^{(k)}$ precum și a valorii funcției într-un punct învecinat, $(X^{(k)} + \lambda^{(k)} \cdot d^{(k)})$.

$$C_0 = F_0 = F_{(\lambda=0)} = F(X^{(k)})$$

$$C_1 = \left. \frac{\partial F}{\partial \lambda} \right|_{\lambda=0} = \lim_{\lambda \rightarrow 0} \frac{\sum_{j=1}^n \frac{\partial F(X)}{\partial X_j}}{\lambda} = \lim_{\lambda \rightarrow 0} \sum_{j=1}^n g_j^{(k)} \left(\frac{\lambda d_j^{(k)}}{\lambda} \right) = (g^{(k)}, d^{(k)})$$

$$F_1 = F(X^{(k+1)}) = F(X^{(k)} + \lambda d^{(k)})$$

Dacă aceste relații se introduc în expresia polinomială a funcției $F(\lambda)$ se va obține:

$$C_2 = \frac{F_1 - \lambda (g^{(k)}, d^{(k)}) - F_0}{\lambda^2}$$

Metoda 2

Constă în folosirea valorii funcției și a derivatei acesteia în punctul curent $X^{(k)}$ precum și a valorii funcției într-un punct învecinat, $(X^{(k)} + \lambda^{(k)} \cdot d^{(k)})$.

$$C_0 = F_0 = F_{(\lambda=0)} = F(X^{(k)})$$

$$C_1 = \left. \frac{\partial F}{\partial \lambda} \right|_{\lambda=0} = \lim_{\lambda \rightarrow 0} \frac{\sum_{j=1}^n \frac{\partial F(X)}{\partial X_j}}{\lambda} = \lim_{\lambda \rightarrow 0} \sum_{j=1}^n g_j^{(k)} \left(\frac{\lambda d_j^{(k)}}{\lambda} \right) = (g^{(k)}, d^{(k)})$$

$$F_1 = F(X^{(k+1)}) = F(X^{(k)} + \lambda d^{(k)})$$

Dacă aceste relații se introduc în expresia polinomială a funcției $F(\lambda)$ se va obține:

$$C_2 = \frac{F_1 - \lambda (g^{(k)}, d^{(k)}) - F_0}{\lambda^2}$$

Metoda 2

Introducând expresiile corespunzătoare coeficienților C_1 și C_2 în relația de minim se obține:

$$\lambda^{(k)} = \frac{1}{2} \cdot \frac{\lambda^2 (g^{(k)}, d^{(k)})}{F_1 - F_0 - \lambda (g^{(k)}, d^{(k)})} = \frac{1}{2} \cdot \frac{\lambda}{1 + \frac{1}{\lambda} \frac{F_0 - F_1}{(g^{(k)}, d^{(k)})}}$$

Metoda 3

Metoda se bazează pe descompunerea în serii Taylor a funcției F , din care se vor reține primii doi termeni:

$$F(X^{(k)} + \lambda d^{(k)}) = F(X^{(k)}) + \lambda (g^{(k)}, d^{(k)}) + 0,5 \lambda^2 (d^{(k)}, H^{(k)} d^{(k)})$$

Metoda 3

Se derivează relația în raport cu λ și se anulează:

$$\lambda^{(k)} = - \frac{\left(g^{(k)}, d^{(k)} \right)}{\left(d^{(k)}, H^{(k)} d^{(k)} \right)}, \quad \left(d^{(k)}, H^{(k)} d^{(k)} \right) > 0$$

Pentru aplicarea acestei metode trebuie să se dispună de valorile derivatei și a hessianului funcției F în punctul $X^{(k)}$.

Metode de explorare directă

Plecând de la un interval inițial $[a_0, b_0]$ care conține valoarea optimă λ^* , metodele de explorare directă reduc iterativ intervalul în care se găsește λ^* până când este satisfăcută o condiție de oprire.

Reducerea se face prin eliminarea simultană sau succesivă a unor subintervale pentru care există certitudinea că nu pot conține valoarea optimă λ^* . Aplicarea reducerii se bazează pe unimodalitatea funcției (are un singur extrem în interval).

Metode pentru rezolvarea problemelor de optimizare convexă fără restricții

Metode de ordinul 0 care folosesc valorile funcției în punctul curent și vecinătăți;

Metode de ordinul 1 care folosesc și derivata de ordinul 1 (gradientul);

Metode de ordinul 2 care folosesc derivatele de ordinul 1 și 2 (hessianul).

Metode de ordinul 0

Avantaje:

- nu sunt necesare expresiile analitice ale derivatei funcției obiectiv $F(X)$.
- nu există condiții referitoare la continuitatea și derivabilitatea funcției $F(X)$.
- printr-o alegere corespunzătoare a criteriilor de stop, se poate evita terminarea iterațiilor într-un punct de inflexiune.

Dezavantaj:

- au o viteză de convergență scăzută.

Metodele de ordinul 0

Au la bază relația iterativă :

$$X^{(k+1)} = X^{(k)} + \lambda^{(k)} d^{(k)}, \quad k = 1, 2, 3, \dots$$

Stabilirea noii direcții de explorare se face pe baza unei serii de evaluări a funcției obiectiv $F(X)$, într-o manieră specifică fiecărei metode, iar lungimea pașilor $\lambda^{(k)}$ este proprie fiecărei metode.

Optimizarea ciclică de-a lungul axelor de coordonate

În cadrul metodei se face o explorare unidimensională pe direcții care coincid cu axele de coordonate.

În relația iterativă direcțiile $d^{(k)}$ sunt succesiv axele sistemului ortogonal de coordonate:

$$d^{(0)} = [1 \ 0 \ 0 \ \dots \ 0]^t, d^{(1)} = [0 \ 1 \ 0 \ \dots \ 0]^t, \dots, d^{(n-1)} = [0 \ 0 \ 0 \ \dots \ 1]^t$$

Pașii $\lambda^{(k)}$ pe direcțiile respective sunt determinați cu una din metodele prezentate.

Procesul iterativ se va încheia când va fi îndeplinit criteriul de stop impus în n iterații.

Dacă suprafețele respective prezintă o vale sau o creastă care nu este paralelă cu axele de coordonate, metoda nu poate fi aplicată.

Exemplu numeric

Să se determine minimul funcției:

$$F(X) = \min(x_1^2 + 4 \cdot x_2^2 - 4)$$

folosind metoda de optimizare ciclică de-a lungul axelor de coordonate, indicându-se ca punct de pornire $X_0 = [5, 4]$.

Iterația 1

Se face deplasarea pe direcția $d^{(0)} = [1 \ 0]^t$, astfel încât pentru determinarea unei noi aproximații se va utiliza relația iterativă:

$$X^{(1)} = X^{(0)} + \lambda^{(0)} d^{(0)} = \begin{bmatrix} 5 \\ 4 \end{bmatrix} + \lambda^{(0)} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 + \lambda^{(0)} \\ 4 \end{bmatrix}$$

Exemplu numeric

Pentru determinarea lungimii pasului de deplasare $\lambda^{(0)}$, se introduc valorile $x_1^{(1)}$ și $x_2^{(1)}$, în expresia funcției obiectiv F :

$$F(\lambda^{(0)}) = (5 + \lambda^{(0)})^2 + 4 \cdot 4^2 - 4 = (\lambda^{(0)})^2 + 10 \cdot \lambda^{(0)} + 85$$

În continuare se face derivata expresiei $F(\lambda^{(0)})$ în raport cu $\lambda^{(0)}$ și se anulează.

$$\frac{\partial F(\lambda^{(0)})}{\partial \lambda^{(0)}} = 2 \cdot \lambda^{(0)} + 10 = 0 \quad \rightarrow \quad \lambda^{(0)} = -5$$

Valoarea pasului de deplasare se introduce în relația iterativă obținându-se astfel valoarea noii aproximații $X^{(1)}$:

$$X^{(1)} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} \end{bmatrix}_t = \begin{bmatrix} 0 & 4 \end{bmatrix}_t$$

Exemplu numeric

Iterația 2

Se face deplasarea pe direcția $d^{(1)} = [0 \ 1]^t$, astfel încât pentru determinarea unei noi aproximații se va utiliza relația iterativă:

$$X^{(2)} = X^{(1)} + \lambda^{(1)}d^{(1)} = \begin{bmatrix} 0 \\ 4 \end{bmatrix} + \lambda^{(1)}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 + \lambda^{(1)} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(1)}$, se introduc valorile $x_1^{(2)}$ și $x_2^{(2)}$, în expresia funcției obiectiv F :

$$F(\lambda^{(1)}) = 0 + 4 \cdot (4 + \lambda^{(1)})^2 - 4 = (\lambda^{(1)})^2 + 8 \cdot \lambda^{(1)} + 15$$

În continuare se face derivata expresiei $F(\lambda^{(1)})$ în raport cu $\lambda^{(1)}$ și se anulează.

$$\frac{\partial F(\lambda^{(1)})}{\delta \lambda^{(1)}} = 2 \cdot \lambda^{(1)} + 8 = 0 \rightarrow \lambda^{(0)} = -4$$

Exemplu numeric

Valoarea pasului de deplasare se introduce în relația iterativă obținându-se astfel valoarea noii aproximații $X^{(2)}$:

$$X^{(2)} = \begin{bmatrix} x_1^{(2)} & x_2^{(2)} \end{bmatrix}^t = \begin{bmatrix} 0 & 0 \end{bmatrix}^t$$

Metoda direcțiilor conjugate

Fie o funcție obiectiv care are o expresie pătratică de forma:

$$F(X) = \frac{1}{2} X^t \cdot C \cdot X + X \cdot D$$

În cadrul acestei metode se pleacă dint-un punct inițial $X^{(0)}$ și se face o deplasare pe direcția $d^{(0)}$, presupusă deocamdată arbitrară,

$$X^{(1)} = X^{(0)} + \lambda^{(0)} d^{(0)}$$

unde pasul $\lambda^{(0)}$ se determină cu una din metodele indicate anterior.

Metoda direcțiilor conjugate

Următoarea direcție de deplasare se va determina din următoarea condiție de conjugare:

$$\begin{aligned} (d^{(1)}, C \cdot d^{(0)}) &= 0 \\ (d^{(i)}, C \cdot d^{(j)}) &= 0, \quad 0 \leq i \neq j \leq n-1 \end{aligned}$$

Folosirea direcțiilor conjugate asigură găsirea minimului unei funcții pătratice de n variabile în n iterații, deplasarea în interiorul fiecărei iterații făcându-se pe o direcție conjugată cu cea din iterația precedentă.

Exemplu numeric

Să se determine minimul funcției:

$$F(X) = \min(x_1^2 + 4 \cdot x_2^2 - 4)$$

folosind metoda direcțiilor conjugate, indicându-se ca punct de pornire $X_0 = [5, 4]$.

Iterația 1

Se face deplasarea pe o direcție arbitrară $d^{(0)} = [d^{(0)}_1 \ d^{(0)}_2]$. Pentru a evidenția acest lucru se consideră o direcție oarecare:

$$d^{(0)} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^t$$

$$|d^{(0)}| = 1$$

Exemplu numeric

Mărimea pasului de deplasare se va face cu ajutorul relației:

$$\lambda^{(k)} = - \frac{\left(g^{(k)}, d^{(k)} \right)}{\left(d^{(k)}, H^{(k)} d^{(k)} \right)}$$

$$g = \left[\frac{\partial F(X)}{\partial X_1} \quad \frac{\partial F(X)}{\partial X_2} \right]_t = [2X_1 \quad 8X_2]^t$$

$$H = C = \begin{bmatrix} \frac{\partial^2 F(X)}{\partial X_1^2} & \frac{\partial^2 F(X)}{\partial X_1 \partial X_2} \\ \frac{\partial^2 F(X)}{\partial X_2 \partial X_1} & \frac{\partial^2 F(X)}{\partial X_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

$$\lambda^{(0)} = - \frac{[10 \quad 32] \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^t}{\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^t} = -5,033$$

Exemplu numeric

Noua aproximație se va calcula utilizând relația iterativă:

$$X^{(1)} = X^{(0)} + \lambda^{(0)} d^{(0)} = \begin{bmatrix} 5 \\ 4 \end{bmatrix} + (-5,033) \cdot \begin{bmatrix} 1 & \sqrt{3} \\ 2 & 2 \end{bmatrix}^t = \begin{bmatrix} 2,484 \\ -0,358 \end{bmatrix}$$

Iterația 2

Se determină direcția de deplasare $d^{(1)}$. Aceasta trebuie să fie conjugată cu direcția $d^{(0)}$, adică trebuie să fie satisfăcută relația:

$$\begin{bmatrix} d_1^{(1)} & d_2^{(1)} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ \sqrt{3} \\ 2 \end{bmatrix} = 0$$

Exemplu numeric

La condiția de conjugare se mai adaugă condiția de vector de modul unitate pentru $d^{(1)}$ dată de relația:

$$\begin{bmatrix} d_1^{(1)} & d_2^{(1)} \end{bmatrix} \begin{bmatrix} d_1^{(1)} & d_2^{(1)} \end{bmatrix}^t = 1$$

Rezolvarea sistemului format din ultimele două ecuații ne va conduce la găsirea direcției de deplasare $d^{(1)} = [d^{(1)}_1 \ d^{(1)}_2]$.

$$d^{(1)} = [-0,9897 \quad 0,1428]^t$$

Pentru determinarea valorii pasului de deplasare $\lambda^{(1)}$ se va calcula gradientul funcției în punctul $X^{(1)}$:

$$g^{(1)} = [2.2,484 \quad -8.0,358]^t = [4,9868 \quad -2,864]^t$$

$$\lambda^{(1)} = - \frac{[4,968 \quad -2,864] [-0,9897 \quad 0,1428]^t}{[-0,9897 \quad 0,1428] \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} -0,9897 \\ 0,1428 \end{bmatrix}^t} = 2,5096$$

Exemplu numeric

Valorile pasului de deplasare și a direcției se vor introduce în relația iterativă obținându-se:

$$X^{(2)} = \begin{bmatrix} 2,484 \\ -0,358 \end{bmatrix} + 2,5096 \begin{bmatrix} -0,9897 \\ 0,1428 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

METODE DE ORDINUL 1

Pentru determinarea direcției și a pasului de deplasare, metodele de ordinul unu apelează la calculul derivatelor de ordinul unu.

pentru o funcție obiectiv $F(X)$ continuă și derivabilă, gradientul într-un punct curent $X^{(k)}$ reprezintă vectorul derivatelor parțiale de ordinul unu în punctul respectiv.

Acest vector este ortogonal la conturul lui $F(X)$ ce trece prin punctul $X^{(k)}$.

Direcția gradientului corespunde celei mai rapide creșteri a lui $F(X)$, ceea ce permite utilizarea ei pentru maximizare sau a direcției opuse, pentru minimizare.

Metoda gradientului simplu

Metoda gradientului simplu este cea mai simplă metodă de ordinul 1 pentru minimizarea unei funcții de mai multe variabile.

Metoda are la bază relația de recurență:

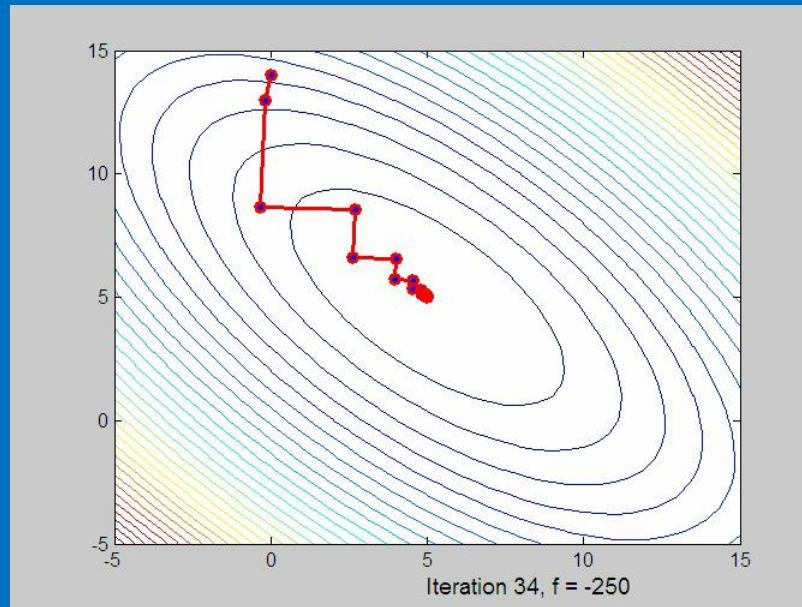
$$X^{(k+1)} = X^{(k)} + \lambda^{(k)} d^{(k)}, \quad k = 1, 2, 3, \dots$$

unde direcția $d^{(k)} = -\nabla F(X^{(k)})$

Avantaj: simplu de aplicat

Dezavantaj: convergența lentă.

Metoda gradientului simplu



- Gradientul este întotdeauna perpendicular pe normala la suprafața funcției obiectiv.
-
- După fiecare iterație noua direcție a gradientului este întotdeauna perpendiculară pe direcția din iterația anterioară.
-
- În consecință, iterațiile au tendința de a asigura o deplasare în zig-zag într-un mod foarte ineficient.

Metoda gradientului conjugat

În cadrul metodei gradientului conjugat noua direcție de deplasare poate fi determinată dacă se utilizează o combinație între gradientul calculat în punctul $X^{(k)}$ și direcția precedentă,

$$d^{(k)} = -g^{(k)} + \beta^{(k)} d^{(k-1)}, k = 1, 2, \dots$$

$$d^{(0)} = -g^{(0)}$$

Există mai multe variante ale metodelor de gradienti conjugați, acestea deosebindu-se prin modul de determinare a parametrului scalar $\beta^{(k)}$:

- metoda Fletcher-Reeves:

$$\beta^{(k)} = \frac{\left(g^{(k)}, g^{(k)} \right)}{\left(g^{(k-1)}, g^{(k-1)} \right)}$$

- metoda Polak-Ribière:

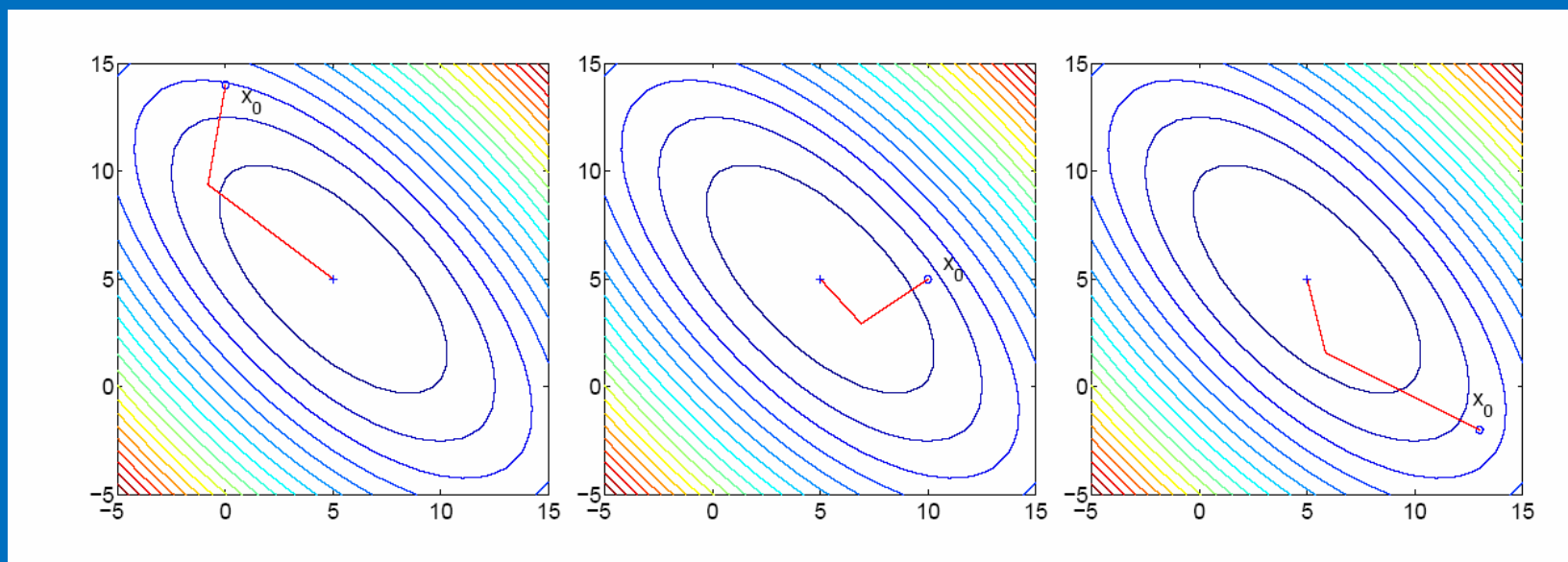
$$\beta^{(k)} = \frac{\left(g^{(k)}, g^{(k)} - g^{(k-1)} \right)}{\left(g^{(k-1)}, g^{(k-1)} \right)}$$

- metoda Hestenes-Stiefel:

$$\beta^{(k)} = \frac{\left(g^{(k)}, g^{(k)} - g^{(k-1)} \right)}{\left(g^{(k)} - g^{(k-1)}, d^{(k-1)} \right)}$$

Metode ale gradinetului conjugat

- Minimul unei funcții pătratice $F(X)$ poate fi determinat în cel mult n iterații, convergența lor fiind bună.



- Se pleacă din 3 puncte diferite;
- Minimul se obține în exact 3 iterații

Exemplu:

Să se minimizeze funcția:

$$\min F(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2x_3 + (x_3)^2 + x_3$$

folosind metoda gradientului simplu.

Se consideră aproximația inițială:

$$X^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Se calculează gradientul și se determină direcția de deplasare:

$$\nabla F(X) = [2x_1 + (1 - x_2) \quad -x_1 + 2x_2 - x_3 \quad -x_2 + 2x_3 + 1]$$

$$\mathbf{d}^0 = -\nabla F(X^0) = -[2(0) + 1 - 0 \quad -0 + 0 - 0 \quad -0 + 0 + 1]$$

$$= -[1 \quad 0 \quad 1] = [-1 \quad 0 \quad -1]$$

$$X^1 = [0 \quad 0 \quad 0] + \lambda^0 \cdot [-1 \quad 0 \quad -1]$$

In continuare se detemină valoarea lui λ^0

Pentru determinarea lungimii pasului de deplasare $\lambda^{(0)}$, se introduc valorile $x_1^{(1)}$ și $x_2^{(1)}$, în expresia funcției obiectiv F :

$$F(X^1) = (\lambda^0)^2 + (-\lambda^0)(1) + 0 - 0 + (\lambda^0)^2 + (-\lambda^0)$$

$$= 2(\lambda^0)^2 - 2(\lambda^0)$$

În continuare se face derivata expresiei $F(\lambda^{(0)})$ în raport cu $\lambda^{(0)}$ și se anulează.

$$\frac{df(X^1)}{d\lambda^0} = 4(\lambda^0) - 2$$

$$4(\lambda^0) = 2 \quad \Rightarrow \quad \lambda^0 = \frac{2}{4} = \frac{1}{2}$$

Valoarea pasului de deplasare se introduce în relația iterativă obținându-se astfel valoarea noii aproximații $X^{(1)}$:

$$X^1 = [0 \ 0 \ 0] + \lambda^0 \cdot [-1 \ 0 \ -1]$$

$$= [0 \ 0 \ 0] + \left[-\frac{1}{2} \ 0 \ -\frac{1}{2} \right]$$

$$X^1 = \left[-\frac{1}{2} \ 0 \ -\frac{1}{2} \right]$$

Se determină noua direcție de deplasare:

$$\mathbf{d}^1 = -\nabla F(X^1) = - \left[-1+1+0 \quad \frac{1}{2}+0+\frac{1}{2} \quad 0-1+1 \right]$$

$$\mathbf{d}^1 = [0 \ -1 \ 0]$$

Se determină noua aproximație:

$$X^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \lambda^1 \cdot [0 \quad -1 \quad 0]$$

$$= \begin{bmatrix} -\frac{1}{2} & -\lambda^1 & -\frac{1}{2} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(1)}$, se introduc valorile $x_1^{(2)}$ și $x_2^{(2)}$, în expresia funcției obiectiv F :

$$F(X^2) = \frac{1}{4} + \left(-\frac{1}{2}\right)(1 + \lambda^1) + (\lambda^1)^2 - (\lambda^1)\left(\frac{1}{2}\right) + \frac{1}{4} - \frac{1}{2}$$

$$= (\lambda^1)^2 - \lambda^1 - \frac{1}{2}$$

În continuare se face derivata expresiei $F(\lambda^{(1)})$ în raport cu $\lambda^{(1)}$ și se anulează.

$$\frac{dF(X^1)}{d\lambda^1} = 2(\lambda^1) - 1$$

$$2(\lambda^1) = 1 \implies$$

$$\lambda^1 = \frac{1}{2}$$

Valoarea pasului de deplasare se introduce în relația iterativă obținându-se astfel valoarea noii aproximații $X^{(2)}$:

$$X^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \lambda^1 \cdot \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$X^2 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Se determină noua direcție de deplasare:

$$\mathbf{d}^2 = -\nabla F(X^2) = -\begin{bmatrix} -1+1+\frac{1}{2} & \frac{1}{2}-1+\frac{1}{2} & \frac{1}{2}-1+1 \end{bmatrix}$$

$$\mathbf{d}^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

Se determină noua aproximație:

$$X^3 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \lambda^2 \cdot \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2}(\lambda^2 + 1) & -\frac{1}{2} & -\frac{1}{2}(\lambda^2 + 1) \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(2)}$, se introduc valorile $x_1^{(3)}$ și $x_2^{(3)}$, în expresia funcției obiectiv F :

$$F(X^3) = \frac{1}{2}(\lambda^2 + 1)^2 - \frac{3}{2}(\lambda^2 + 1) + \frac{1}{4}$$

În continuare se face derivata expresiei $F(\lambda^{(2)})$ în raport cu $\lambda^{(2)}$ și se anulează.

$$\frac{dF(X^3)}{d\lambda^2} = (\lambda^2 + 1) - \frac{3}{2}$$

$$(\lambda^2 + 1) = \frac{3}{2} \Rightarrow \lambda^2 = \frac{1}{2}$$

Se determină noua aproximație:

$$X^3 = \begin{bmatrix} 1 & 1 & 1 \\ -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \lambda^2 \cdot \begin{bmatrix} 1 & 0 & 1 \\ -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 \\ -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ -\frac{1}{4} & 0 & -\frac{1}{4} \end{bmatrix}$$

$$X^3 = \begin{bmatrix} 3 & 1 & 3 \\ -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix}$$

Se determină noua direcție de deplasare:

$$\mathbf{d}^3 = -\nabla F(X^3) = -\begin{bmatrix} 0 & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$X^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix} + \lambda^3 \cdot \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2}(\lambda^3 + 1) & -\frac{3}{4} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(3)}$, se introduc valorile $x_1^{(4)}$ și $x_2^{(4)}$, în expresia funcției obiectiv F :

$$F(X^4) = \frac{1}{4}(\lambda^3 + 1)^2 - \frac{3}{2}(\lambda^3) - \frac{3}{2}$$

$$\frac{dF(X^4)}{d\lambda^3} = \frac{1}{2}(\lambda^3 + 1) - \frac{9}{8} = 0$$

$$\Rightarrow \lambda^3 = \frac{5}{4}$$

Se determină noua aproximație:

$$X^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix} + \begin{bmatrix} 0 & -\frac{5}{8} & 0 \end{bmatrix}$$

$$X^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix}$$

Se determină noua direcție de deplasare:

$$\mathbf{d}^4 = -\nabla F(X^4) = -\begin{bmatrix} \frac{5}{8} & -\frac{3}{4} & \frac{5}{8} \\ \frac{5}{8} & -\frac{3}{4} & \frac{5}{8} \end{bmatrix} = \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \\ -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

Se determină noua aproximație:

$$X^5 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \\ -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix} + \lambda^4 \cdot \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \\ -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{4} \left(3 + \frac{5}{2} \lambda^4\right) & -\frac{3}{4} \left(\frac{3}{2} - \lambda^4\right) & -\frac{1}{4} \left(3 + \frac{5}{2} \lambda^4\right) \\ -\frac{1}{4} \left(3 + \frac{5}{2} \lambda^4\right) & -\frac{3}{4} \left(\frac{3}{2} - \lambda^4\right) & -\frac{1}{4} \left(3 + \frac{5}{2} \lambda^4\right) \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(4)}$, se introduc valorile $\mathbf{x}_1^{(5)}$ și $\mathbf{x}_2^{(5)}$, în expresia funcției obiectiv F :

$$F(\mathbf{X}^5) = \frac{73}{32} (\lambda^4)^2 - \frac{43}{32} \lambda^4 - \frac{51}{64}$$

$$\frac{dF(\mathbf{x}^5)}{d\lambda^4} = \frac{73}{16} \lambda^4 - \frac{43}{32} = 0 \quad \Rightarrow \quad \lambda^4 = \frac{43}{146}$$

Se determină noua aproximație:

$$\mathbf{X}^5 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix} + \frac{43}{146} \cdot \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

$$\mathbf{X}^5 = \begin{bmatrix} -\frac{1091}{1168} & -\frac{66}{73} & -\frac{1091}{1168} \end{bmatrix}$$

Se verifică dacă criteriul de convergență este satisfăcut:

$$||\nabla F(\mathbf{X}^5)||:$$

$$\nabla F(\mathbf{X}^5) = \begin{bmatrix} \frac{21}{584} & \frac{35}{584} & \frac{21}{584} \end{bmatrix}$$

$$\|\nabla F(\mathbf{X}^5)\| = \sqrt{\left(\frac{21}{584}\right)^2 + \left(\frac{35}{584}\right)^2 + \left(\frac{21}{584}\right)^2} = 0.0786$$

$\|\nabla F(\mathbf{X}^5)\| = 0.0786$, are o valoare foarte mică, apropiată de 0, astfel încât:

$$\mathbf{X} = \begin{bmatrix} -\frac{1091}{1168} & -\frac{66}{73} & -\frac{1091}{1168} \end{bmatrix}$$

este foarte aproape de valoare de optim, obținută analitic:

$$\mathbf{x}^* = \begin{bmatrix} -1 & -1 & -1 \end{bmatrix}$$

METODE DE ORDINUL 2

Metode de tip Newton

Utilizarea gradientului ca direcție de explorare este conformă cu aproximarea liniară a funcției obiectiv în dezvoltarea acesteia în serie Taylor. Practic se poate adopta o aproximare pătratică de forma:

$$F(X) = F(X^{(k)}) + (g^{(k)}, \Delta X^{(k)}) + \frac{1}{2} (d^{(k)}, H^{(k)} d^{(k)})$$

Minimul se obține prin derivarea expresiei în raport cu componentele lui $d^{(k)}$ și anularea derivatelor respective:

$$d^{(k)} = -H^{(k)^{-1}} g^{(k)}$$

Relația care stă la baza determinării direcției se introduce în formula iterativă rezultând:

$$X^{(k+1)} = X^{(k)} - \lambda^{(k)} \cdot H^{(k)^{-1}} g^{(k)}$$

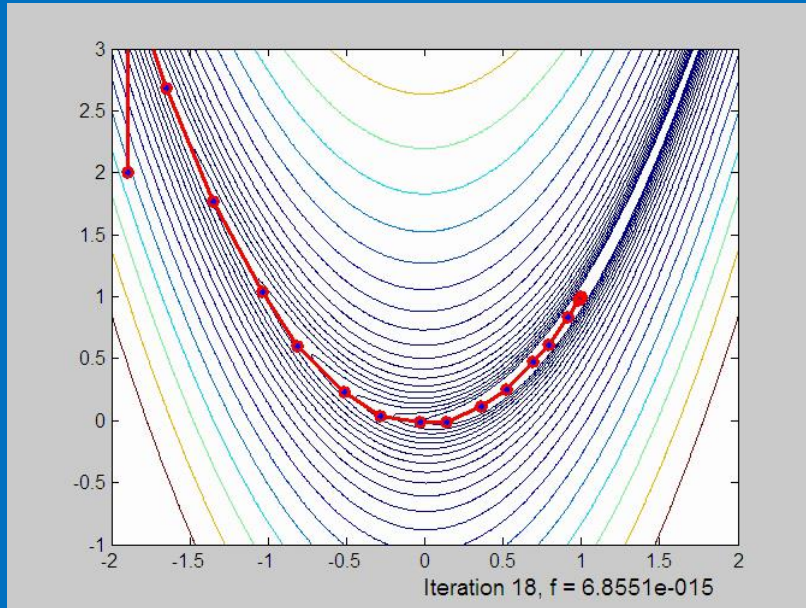
Metoda Newton

În cazul unor funcții obiectiv pătratice, minimumul va fi atins încă din prima iterație, pentru $\lambda^{(0)} = 1$.

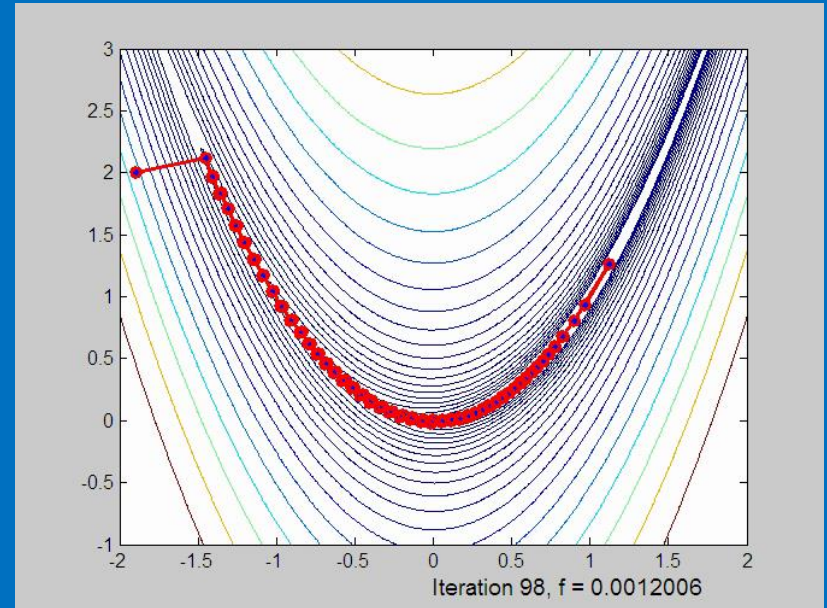
Observații:

- Aplicare metodei Newton cere ca matricea $H^{(k)}$ să fie pozitiv definită la fiecare iterație.
- Convergența metodei în apropierea optimului va fi sensibil mai bună decât a metodelor de ordinul 1.
- disponibilitatea derivatelor de ordinul doi permite verificarea condițiilor de suficiență a optimului.

Metoda Newton



Metoda Newton – 18 iterații



Metoda gradientului conjugat – 98 iterații

Metoda Newton

Dezavantaje:

- volumului de calcul necesar obținerii derivatelor de ordinul doi și pentru inversarea matricei $H^{(k)}$.
- din cauza condiției $H^{(k)} > 0$ pot apărea situații în care metoda nu poate fi aplicată (cazurile în care $F(X)$ este liniară pe anumite porțiuni, iar $H^{(k)} = 0$).
- pot apărea unele cazuri în care $F(X^{(k+1)}) > F(X^{(k)})$, deși $H^{(k)} > 0$, aproximația pătratică fiind valabilă doar într-o zonă foarte limitată.

Metoda Newton

Exemplul 1.

Să se determine minimul funcției:

$$F(X) = \min(x_1^2 + 4 \cdot x_2^2 - 4)$$

folosind metoda Newton, indicându-se ca punct de pornire $X^{(0)} = [5, 4]$, iar $\lambda^{(0)} = 1$.

Se calculează gradientul și hesianul funcției $F(X)$:

$$g = \left[\frac{\partial F(X)}{\partial x_1} \quad \frac{\partial F(X)}{\partial x_2} \right]^t = [2x_1 \quad 8x_2]^t$$

$$H = C = \begin{bmatrix} \frac{\partial^2 F(X)}{\partial x_1^2} & \frac{\partial^2 F(X)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 F(X)}{\partial x_2 \partial x_1} & \frac{\partial^2 F(X)}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

$$X^{(1)} = X^{(0)} - H^{(0)^{-1}} \cdot g^{(0)} = \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 32 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Exemplul 2.

Să se determine minimul funcției:

$$f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = (\mathbf{x}_1)^2 + \mathbf{x}_1(1 - \mathbf{x}_2) + (\mathbf{x}_2)^2 - \mathbf{x}_2\mathbf{x}_3 + (\mathbf{x}_3)^2 + \mathbf{x}_3$$

Se calculează gradientul:

$$\nabla f(x_1, x_2, x_3) = [2x_1 - x_2 + 1 \quad -x_1 + 2x_2 - x_3 \quad -x_2 + 2x_3 + 1]$$

Matricea Hessiană:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Matricea Hessiană inversată:

$$\left[\nabla^2 f(\mathbf{x}) \right]^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{bmatrix}$$

Se alege punctul inițial de pornire:

$$\mathbf{x}^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Se calculează valoarea gradientului în punctul curent:

$$\nabla f(\mathbf{x}^0) = [0 - 0 + 1 \quad -0 + 0 - 0 \quad -0 + 0 + 1]$$

$$\Rightarrow \nabla f(\mathbf{x}^0) = [1 \quad 0 \quad 1]$$

Se scrie relația iterativă pentru determinarea
noii aproximații:

$$\begin{aligned}\mathbf{x}^1 &= \mathbf{x}^0 - \left[\nabla^2 f(\mathbf{x}^0) \right]^{-1} \cdot \nabla f(\mathbf{x}^0) \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\end{aligned}$$

$$\therefore \mathbf{x}^1 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Se calculează gradientul funcției în noua aproximație:

$$\nabla f(\mathbf{x}^1) = [-2+1+1 \quad 1-2+1 \quad 1-2+1] = [0 \quad 0 \quad 0]$$

Deoarece valoarea gradientului este 0, metoda a fost convergentă.

Metode cvasiNewton

Acestea din urmă au fost elaborate în scopul evitării principalelor dezavantaje ale metodelor Newton referitoare la inversarea matricei hessiene în fiecare iterație.

Metodele quasi-Newton au la bază relația de recurență,

$$X^{(k+1)} = X^{(k)} - \lambda^k \cdot \tilde{H}^{(k)} g^{(k)}$$

unde matricea $\tilde{H}^{(k)}$ reprezintă o aproximație lui $H^{(k)-1}$

Relația iterativă de mai sus poate fi considerată ca fiind relația fundamentală care stă la baza metodelor ce utilizează derivatele.

Cel mai performant algoritm din această categorie de metode este *algoritmul Davidson-Fletcher-Powell*.

Algoritmul Davidson-Fletcher-Powell

Caracteristic acestui algoritm este faptul că modificarea matricei $\tilde{H}^{(k)}$ la fiecare iterație este făcută astfel încât pentru o funcție pătratică de n variabile, în limita a n iterații, matricea $\tilde{H}^{(k)}$ să devină egală cu matricea $H^{(k)-1}$.

Inițial, matricea $\tilde{H}^{(k)}$ se alege ca fiind matricea unitate I_n , astfel încât prima iterație decurge după metoda gradientului. Ulterior se face o modificare treptată de la direcția gradientului la cea a metodei Newton, reușindu-se astfel combinarea avantajelor aferente celor două metode.

$$\tilde{H}^{(k+1)} = \tilde{H}^{(k)} + A^{(k)} - B^{(k)} = \tilde{H}^{(k)} + \frac{[\Delta X^{(k)}][\Delta X^{(k)}]^t}{(\Delta X^{(k)}, \Delta g^{(k)})} - \frac{[\tilde{H}^{(k)}][\Delta g^{(k)}][\Delta g^{(k)}]^t [\tilde{H}^{(k)}]^t}{(\Delta g^{(k)}, H^{(k)} \Delta g^{(k)})}$$

Algoritmul Davidson-Fletcher-Powell

unde:

$$\Delta X^{(k)} = X^{(k+1)} - X^{(k)}$$

$$\Delta g^{(k)} = g^{(k+1)} - g^{(k)}$$

Rolul matricei $A^{(k)}$ este de a asigura convergența lui $\tilde{H}^{(k)}$ către $H^{(k)-1}$, în timp ce matricea $B^{(k)}$ are rolul anulării treptate a efectului alegerii inițiale, adică a matricei unitate.

Pentru cazul funcțiilor obiectiv pătratice direcțiile utilizate de algoritmul Davidson-Fletcher-Powell sunt direcții conjugate.

Algoritmul Davidson-Fletcher-Powell

Algoritmul evoluează satisfăcător dacă se îndeplinesc următoarele condiții:

- eroarea în evaluarea componentelor gradientului $g^{(k)}$ nu este prea mare (în cazul derivării numerice);
- Matricea $\tilde{H}^{(k)}$ nu este singulară. În situația în care aceasta este aproape de singularitate una din soluțiile propuse constă în reinițializarea ei cu matricea unitate (există și alte posibilitați).

Algoritmul Davidson-Fletcher-Powell

Exemplu numeric:

Să se determine minimul funcției:

$$F(X) = \min \left(4 \cdot (x_1 - 2)^2 + (x_2 - 3)^2 \right)$$

folosind metoda Davidson-Fletcher-Powell, indicându-se ca punct de pornire $X^{(0)} = [6, 5]$.

Se calculează gradientul iar hesianul funcției este matricea unitate :

$$g^{(k)} = \begin{bmatrix} 8(x_1 - 2) \\ 2(x_2 - 3) \end{bmatrix}_{X=X^{(k)}} = \begin{bmatrix} 32 \\ 4 \end{bmatrix}$$

$$X^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \end{bmatrix} - \lambda^{(0)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 32 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 - 32\lambda^{(0)} \\ 5 - 4\lambda^{(0)} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(0)}$, se introduc valorile $x_1^{(1)}$ și $x_2^{(1)}$ în expresia funcției obiectiv F , se face derivata în raport cu $\lambda^{(0)}$, după care se anulează:

$$\frac{\partial F(x^{(1)})}{\partial \lambda^{(0)}} = 65 - 514 \lambda^{(0)} = 0$$

$$\lambda^{(0)} = 0,1264$$

Valoarea noii aproximații este:

$$X^{(1)} = \begin{bmatrix} 6 \\ 5 \end{bmatrix} - 0,1264 \cdot \begin{bmatrix} 32 \\ 4 \end{bmatrix} = \begin{bmatrix} 1,955 \\ 4,494 \end{bmatrix}$$

Iterația 2

$$g^{(1)} = \begin{bmatrix} -0,36 \\ 2,988 \end{bmatrix}$$

$$\Delta X^{(1)} = \begin{bmatrix} -4,045 \\ -0,506 \end{bmatrix}$$

$$\Delta g^{(1)} = \begin{bmatrix} -32,36 \\ -1,012 \end{bmatrix}$$

Se calculează matricea $\tilde{H}^{(1)}$:

$$\tilde{H}^{(k+1)} = \tilde{H}^{(k)} + A^{(k)} - B^{(k)} = \tilde{H}^{(k)} + \frac{[\Delta X^{(k)}][\Delta X^{(k)}]^t}{(\Delta X^{(k)}, \Delta g^{(k)})} - \frac{[\tilde{H}^{(k)}][\Delta g^{(k)}][\Delta g^{(k)}]^t [\tilde{H}^{(k)}]^t}{(\Delta g^{(k)}, H^{(k)} \Delta g^{(k)})}$$

$$H^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\begin{bmatrix} -4,045 & 0 \\ -0,506 & 0 \end{bmatrix} \cdot \begin{bmatrix} -4,045 & -0,506 \\ 0 & 0 \end{bmatrix}}{\begin{bmatrix} -4,045 & -0,506 \end{bmatrix} \cdot \begin{bmatrix} -32,36 \\ -1,012 \end{bmatrix}} + \frac{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -32,36 & 0 \\ -1,012 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}{\begin{bmatrix} -32,36 & -1,012 \end{bmatrix} \cdot \begin{bmatrix} -32,36 \\ -1,012 \end{bmatrix}} = \begin{bmatrix} 0,1554 & -0,0147 \\ -0,0147 & 1,001 \end{bmatrix}$$

În continuare se poate calcula noua aproximație $X^{(2)}$:

$$X^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \end{bmatrix} = \begin{bmatrix} 1,955 \\ 4,494 \end{bmatrix} - \lambda^{(1)} \begin{bmatrix} 0,1554 & -0,0147 \\ -0,0147 & 1,0010 \end{bmatrix} \cdot \begin{bmatrix} -0,36 \\ 4,494 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

unde $\lambda^{(1)} = 0,4984$ a fost determinat în mod similar ca $\lambda^{(0)}$.

Deoarece:

$$g^{(2)} = [0 \quad 0]^t$$

calculele sunt terminate.

PROGRAMAREA NELINIARĂ CU RESTRICȚII

GENERALITĂȚI

Soluția optimă trebuie să fie aleasă în mod obligatoriu în interiorul domeniului admisibil pentru vectorul variabilelor de optimizare $X = [x_1, x_2, \dots, x_n]$.

Problema care trebuie rezolvată se referă la determinarea valorii variabilelor care asigură extremul unei funcții obiectiv:

$$\min F(X) = \min F(x_1, x_2, \dots, x_n)$$

în prezența restricțiilor:

$$g_i(x_1, x_2, \dots, x_n) \leq 0, \quad i = 1, \dots, m$$

$$x_j \geq 0, \quad j = 1, \dots, n$$

Condițiile Kuhn-Tucker (K-T)

Condițiile care trebuie să fie satisfăcute pentru un optim cu restricții, local sau global au fost determinate de K-T având ca punct de plecare modelele clasice ale multiplicatorilor lui Lagrange.

A. Condițiile K-T pentru cazul $g_i(X)$

Funcția Lagrange asociată problemei PN cu restricții (prezentată anterior), în care $F(X)$ și $g_i(X)$ au derivate parțiale de ordinul întâi, în raport cu x_j , $j = 1, \dots, n$, respectiv $g_i(X)$ sunt doar restricții de egalitate, are următoarea formă:

$$\Phi(x, \lambda) = F(x) + \sum_{i=1}^m \lambda_i g_i(X)$$

unde λ_i , $i = 1, \dots, m$, sunt multiplicatorii Lagrange.

Condițiile Kuhn-Tucker (K-T)

Un punct extrem al funcției $\Phi(X, \lambda)$, care poate fi punctul de minim, se poate obține prin anularea derivatelor de ordinul 1:

$$\frac{\partial \Phi(X, \lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1, \dots, n$$

$$\frac{\partial \Phi(X, \lambda)}{\partial \lambda_i} = g_i(X) = 0, \quad i = 1, \dots, m$$

Relațiile reprezintă condițiile necesare dar nu și suficiente pentru minim. Dacă funcția obiectiv $F(X)$ este **convexă**, iar **restricțiile sunt liniare** atunci este asigurată și **suficiența**.

Concluzie:

Minimul funcției $F(X)$, în prezența restricțiilor de egalitate $g_i(X) = 0$, este obținut pentru acele valori ale lui X care minimizează fără restricții funcția auxiliară $\Phi(X, \lambda)$.

Condițiile Kuhn-Tucker (K-T)

ALGORITM:

1. Se aleg inițial valori pentru λ_i , $i = 1, \dots, m$;
2. Se rezolvă sistemul

$$\frac{\partial \Phi(X, \lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1, \dots, n$$

de n ecuații cu n necunoscute.

3. Soluțiile determinate care depind de valorile lui λ_i sunt introduse în cele m restricții.
4. Dacă restricțiile nu sunt satisfăcute, se aleg alte valori pentru λ_i , $i = 1, \dots, m$, până când cele $(n+m)$ ecuații vor fi satisfăcute de valorile alese.

Condițiile Kuhn-Tucker (K-T)

B. Condițiile K-T pentru cazul $g_i(X) \leq 0$

Utilizând variabile auxiliare, w_i , $i = 1, \dots, m$, inegalitățile se pot transforma în egalități:

$$g_i(X) + w_i^2 = 0, \quad i = 1, \dots, m$$

Funcția Lagrange asociată problemei PN cu restricții are forma:

$$\Phi(X, w, \lambda) = F(X) + \sum_{i=1}^m \lambda_i (w_i^2 + g_i(X))$$

care conduce la un punct extrem dacă sunt îndeplinite condițiile:

Condițiile Kuhn-Tucker (K-T)

$$\frac{\partial \Phi(X, w, \lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1, \dots, n$$

$$\frac{\partial \Phi(X, w, \lambda)}{\partial \lambda_i} = g_i(X) + w_i^2 = 0, \quad i = 1, \dots, m$$

$$\frac{\partial \Phi(X, w, \lambda)}{\partial w_i} = 2\lambda_i w_i = 0, \quad i = 1, \dots, m$$

Deoarece mărimile w_i sunt variabile auxiliare, acestea pot fi eliminate.

$g_i(X) = 0$, independent de i , numai dacă $w_i = 0$.

A treia relație poate fi rescrisă, ținând cont de relația a doua:

$$\lambda_i g_i(X) = 0, \quad i = 1, \dots, m$$

Condițiile Kuhn-Tucker (K-T)

$$\frac{\partial \Phi(X, w, \lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1, \dots, n$$

$$g_i(X) \leq 0, \quad i = 1, \dots, m$$

$$\lambda_i g_i(X) = 0, \quad i = 1, \dots, m$$

$$\lambda_i \geq 0, \quad i = 1, \dots, m$$

condițiile Kuhn-Tucker

Condițiile necesare pentru un punct de minim local și global.

În situația în care $F(X)$ și $g_i(X)$ sunt convexe, relațiile ne dau minimul global.

Condițiile Kuhn-Tucker (K-T)

Observații

Semnul multiplicatorilor λ_i , $i = 1, \dots, m$, se schimbă în funcție de natura optimului căutat și de semnul restricțiilor.

Natura restricțiilor	Natura optimului	
	max $F(X)$	min $F(X)$
$g_i(X) \leq 0, i = 1, \dots, m$	$\lambda_i \leq 0$	$\lambda_i \geq 0$
$g_i(X) \geq 0, i = 1, \dots, m$	$\lambda_i \geq 0$	$\lambda_i \leq 0$

Condițiile Kuhn-Tucker ne indică faptul că pentru o restricție care nu este activă, de tipul:

$$\sum_{j=1}^n a_{ij} \cdot x_j = b_i$$

multiplicatorul $\lambda_i = 0$.

Metode ale programării neliniare cu restricții

Metodele programării neliniare cu restricții se împart în două mari categorii:

- metode ale direcțiilor admisibile;
- metode ale funcțiilor penalizare.

A. Metode ale direcțiilor admisibile

Sunt concepute astfel încât să folosească tehnici de minimizare fără restricții, dar în același timp să considere și restricțiile de inegalitate.

În acest sens se alege un punct inițial $X^{(0)}$ și o direcție astfel încât o deplasare mică de-a lungul ei să nu modifice restricțiile, iar funcția obiectiv $F(X)$ să se îmbunătățească.

Metode ale programării neliniare cu restricții

Direcția de deplasare care satisface prima cerință se numește *direcția admisibilă*.

Astfel, plecând din punctul inițial $X^{(0)}$, pe direcția $d^{(0)} = -\nabla F(X^{(0)})$, se obține $X^{(1)}$.

Procedeul se repetă până în momentul în care este imposibil să se mai găsească o direcție admisibilă.

Metode ale programării neliniare cu restricții

B. Metode ale direcțiilor admisibile

Problema minimizării funcției $F(X)$ în prezența restricțiilor $g_i(X) \geq 0$, $i = 1, 2, \dots, m$, poate fi abordată prin definirea unei funcții de penalizare de tipul:

$$\varphi(y) = \begin{cases} 0, & y \geq 0 \\ \infty, & y < 0 \end{cases}$$

și considerarea minimizării funcției fără restricții:

$$\psi(X) = F(X) + \sum_{i=1}^m \varphi(g_i(X))$$

Metode ale programării neliniare cu restricții

Observații

Dacă toate restricțiile $g_i(X) \geq 0$ sunt satisfăcute, al doilea termen nu contribuie cu nimic la valoarea lui $\Psi(X)$.

Dacă avem o restricție de tipul $g_i(X) < 0$, atunci $\varphi(g_i) = \infty$ și vom fi departe de minim.

Al doilea termen din suma respectivă penalizează orice încălcare a restricțiilor.

Metode ale programării neliniare cu restricții

Dificultăți în abordare

Există situații pentru care trebuie aleasă o funcție de penalizare mai puțin rigidă, care să prezinte proprietățile de regularitate necesară. De exemplu, atunci când funcția obiectiv prezintă discontinuități.

O alegere a unei funcții de tipul:

$$\varphi(y) = \begin{cases} 0, & y \geq 0 \\ y^2 & y < 0 \end{cases}$$

este mai potrivită. Ea este continuă și are derivată continuă pentru toate valorile lui y . Această metodă nu este foarte eficientă în toate cazurile.

Metode ale programării neliniare cu restricții

Metoda Fiacco-McCormick

Pentru a evita dificultățile întâlnite anterior, metoda determină punctul de minim optimul dinspre interiorul mulțimii restricțiilor. Se definește funcția,

$$Z(X, r) = F(X) + r \cdot \sum_{i=1}^m \frac{1}{g_i(X)}, \quad r > 0$$

Minimul acestei funcții trebuie să se realizeze în interiorul mulțimii de restricții deoarece pe această frontieră este îndeplinită condiția:

$$Z(X, r_0) \rightarrow +\infty \quad (g_i(X) = 0)$$

Punctul de minim depinde de r_0 și va fi notat $X(r_0)$. Metoda prezintă interes mai ales atunci când avem de considerat restricții puternic neliniare, eliminând complet mișcarea de-a lungul frontierelor mulțimii restricțiilor.

Metode ale programării neliniare cu restricții

ALGORITM

1. Se alege $r_0 > 0$ și $X^{(0)}$ în interiorul mulțimii de restricții.
Se consideră problema minimizării fără restricții a funcției $Z(X, r)$.
 2. Se repetă acest proces de minimizare pentru un șir de valori $\{r\} = \{r_0 > r_1 > r_2 > \dots > r_k > 0\}$.
Prin micșorarea lui r , influența termenului al doilea, care penalizează apropierea de frontiera mulțimii restricțiilor, se reduce și efortul în minimizarea a funcției $Z(X, r)$ se concentrează asupra lui $F(X)$.
- Șirul $\{X(r_0), X(r_1), \dots\}$ se poate apropia din ce în ce mai mult de frontiera mulțimii restricțiilor, dacă aceasta este avantajos pentru micșorarea funcției $F(X)$.